DOCUMENT RESUME

ED 229 370 SP 022 248

AUTHOR Hord, Shirley M., Ed.; And Others
TITLE Toward Usable Strategies for Teacher Education Program Evaluation. Teacher Education Networks.
INSTITUTION Texas Univ., Austin. Research and Development Center for Teacher Education.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.; Texas A and M Univ., College Station. Coll. of Education.
REPORT NO RDCTE-7010
PUB DATE Apr 82
NOTE 178p.
PUB TYPE Viewpoints (120) -- Information Analyses (070) -- Collected Works - General (020)

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS Accreditation (Institutions); Beginning Teachers; *Evaluation Criteria; *Evaluation Methods; *Evaluation Needs; Followup Studies; Futures (of Society); Higher Education; Minimum Competencies; Needs Assessment; Program Development; Program Effectiveness; *Program Evaluation; Self Evaluation (Groups); *Teacher Education Programs

ABSTRACT

Papers by teacher education practitioners, on the topic of evaluation of teacher education programs, with responses by evaluation experts, are presented: (1) "Teacher Education's Evaluation of Graduates: Where Are We Going and How Do We Know When We Get There?" (J. T. Sandefur); (2) "Requirements of a Data Base for Effective Program Evaluation" (Robert A. Roth); (3) "A Practitioner's Perceptions Regarding Problems in Assessing the Effectiveness of Teacher Education Programs" (Maurice C. Erly); (4) "NCATE Accreditation: Problems, Issues, and Needed Research" (William E. Gardner); (5) "Program Evaluation in Teacher Education: From Admissions through Follow-Up" (Gary R. Galluzzo); (6) "Building Program Ownership: A Collaborative Approach to Defining and Evaluating the Teacher Training Program" (Gary D. Borich); (7) "Introductory Guidelines for Designing Evaluations of Teacher Education Programs" (Nick L. Smith); (8) "Explorations in the Evaluation of Teacher Education" (Daniel L. Stufflebeam); and (9) "Program Evaluation in Teacher Education: Future Directions" (Marlene I. Strathe). (CJ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED229370

TEACHER EDUCATION NETWORKS

SP

rd
te

TOWARD USABLE STRATEGIES FOR TEACHER EDUCATION PROGRAM EVALUATION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
O. Bown
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

editors
Shirley M. Hord
Tom V. Savage
Lowell J. Bethel

SP 022 248

TOWARD USABLE STRATEGIES FOR TEACHER EDUCATION PROGRAM EVALUATION

editors

Shirley M. Hord
Tom V. Savage
Lowell J. Bethel

Supported by

College of Education, Texas A&M University
College of Education, University of Texas, Austin
Research & Development Center for Teacher Education, Austin

1982

The research described herein was conducted under contract with the National Institute of Education. The opinions expressed are those of the authors and do not necessarily reflect the position or policy of the National Institute of Education, and no endorsement by the National Institute of Education should be inferred.

Publication #7010 available from
the Research and Development Center for Teacher Education.

TABLE OF CONTENTS

Page

INTRODUCTION

Shirley Hord, Research and Development Center for Teacher Education, The University of Texas at Austin

Tom Savage, College of Education, Texas A&M University

Lowell Bethel, College of Education, The University of Texas at Austin 1

PRACTITIONERS

TEACHER EDUCATION'S EVALUATION OF GRADUATES: WHERE ARE WE GOING AND HOW DO WE KNOW WHEN WE GET THERE? J. T. Sandefur, Western Kentucky University 9

REQUIREMENTS OF A DATA BASE FOR EFFECTIVE PROGRAM EVALUATION, Robert A. Roth, Michigan State Department of Education 23

A PRACTITIONER'S PERCEPTIONS REGARDING PROBLEMS IN ASSESSING THE EFFECTIVENESS OF TEACHER EDUCATION PROGRAMS, Maurice C. Erly, Prince George's County Public Schools 39

NCATE ACCREDITATION: PROBLEMS, ISSUES, AND NEEDED RESEARCH, William E. Gardner, University of Minnesota . . . 51

PROGRAM EVALUATION IN TEACHER EDUCATION: FROM ADMISSIONS THROUGH FOLLOW-UP, Gary R. Galluzzo, Glassboro State College 67

EVALUATORS

BUILDING PROGRAM OWNERSHIP: A COLLABORATIVE APPROACH TO DEFINING AND EVALUATING THE TEACHER TRAINING PROGRAM, Gary D. Borich, The University of Texas at Austin 81

INTRODUCTORY GUIDELINES FOR DESIGNING EVALUATIONS OF TEACHER EDUCATION PROGRAMS, Nick L. Smith, Northwest Regional Educational Laboratory 101

EXPLORATIONS IN THE EVALUATION OF TEACHER EDUCATION, Daniel L. Stufflebeam, Western Michigan University 131

RESPONSE

PROGRAM EVALUATION IN TEACHER EDUCATION: FUTURE DIRECTIONS, Marlene I. Strathe, Educational Research and Development Center, University of Northern Iowa 169

APPENDIX

LIST OF PARTICIPANTS
April, 1982
AUSTIN CONFERENCE 179

INTRODUCTION

What procedures can be used for assessing the effectiveness of teacher preparation programs and for obtaining data on which to base program improvement? One teacher education task force studying evaluation and program development assessed the state of the art of evaluation in teacher education as "one of conflicting research results, lack of clarity and agreement about evaluation designs, and a fragmentation of evaluation focus" (Strathe, 1982). Such evaluation of teacher preparation programs has been primarily focused on follow-up studies of graduates. The studies have examined various information about graduates including graduates' perceptions about and satisfaction with their programs, and ratings of the graduates' performance by supervising teachers, principals and others. Despite the pressure exerted by internal needs for evaluation and external mandates from the state or national level for accreditation, the state of the art has advanced little. Cooper (1980) observes that "very little of what is known about evaluation models and assessment procedures have been applied to teacher education programs. The methodologies ... are not well developed."

To help resolve this dilemma a collaborative effort was designed and organized by three institutions: The College of Education, Texas A&M University, College Station; The College of Education, The University of Texas, Austin; and, the Research and Development Center for Teacher Education, The University of Texas, Austin. The purpose of the endeavor was to study the lack of evaluation tools for program evaluation and development, to identify specifications for alternative evaluation models or tools, and to direct evaluation experts in the conceptualization and development of such alternatives. The process for reaching the goals of this developmental work covered three phases: needs assessing, diagnosis, and prescription.

Needs Assessing

Five teacher educator practitioners were invited to reflect on their experience with program evaluation, to examine studies of program

evaluation and the research base on it, and to provide recommendations concerning unmet needs and challenges. In brief, they were asked to raise issues requiring attention.

Diagnosis

The written remarks of the practitioners were shared at a national invitational working conference. (The complete list of participants is included in the Appendix.) The practitioners' recommendations provided the stimulus for reaction, interaction and additional input from all conference participants and from three evaluation "experts" who were in attendance in a consultation role. In combination, they articulated clear expressions of reality-based assignments to the evaluators.

Prescription

Each of the three evaluators attended to the practitioners' remarks and developed a paper designed to respond to the common concerns and crucial problems identified by the teacher preparation program evaluation users.

The products of the work just described are contained in this volume. The first part includes the five papers prepared by the practitioners to represent the state of the art, as they view it, and to declare their needs. Following those papers are the three papers from the evaluators which provide concepts, methods, strategies, tools, procedures... to respond to the practitioners' needs. A last piece provides reflections and concluding comments about the entire set of eight papers.

A brief guide to the papers follows.

The Practitioners

J. T. Sandefur, Dean, College of Education, Western Kentucky University, in his paper, "Teacher Education's Evaluation of Graduates:

Where Are We Going and How Do We Know When We Get There?" reflects on his years of experience in program evaluation. He expresses concern about the lack of the impact of evaluation data obtained from evaluation studies -- impact on the development and revision of the ongoing teacher education program. Sandefur reports on the progress that has been made and proposes two major questions for consideration in an evaluation program, "what do we evaluate?" and "how do we evaluate it?" He then divides those two broad topics into "subcomponents" which need to be dealt with. In conclusion, he expresses a sense of urgency to move meaningfully with the implementation of effective evaluation systems which are tied directly into program improvement.

In "Requirements of a Data Base for Effective Program Evaluation," Robert A. Roth, Director of Teacher Preparation and Certification Services, Michigan State Department of Education, discusses a number of issues in need of consideration in this domain. He examines the need for evaluation and the need for a data base for decision-making. Purposes for doing evaluation and the problems encountered with accreditation and program approval are enumerated. NCATE and its relevance to evaluating programs is commented on. Roth discusses the need to focus on the quality of graduates as the outcome of teacher education programs. He provides suggestions for doing this in a structured, comprehensive way and concludes by citing criteria for use in the design of a "program evaluation/approval model."

Maurice C. Erly who is school-based, as contrasted with institution of higher education-based, provided the third paper. Erly, the Coordinating Supervisor of Staff Development in Upper Marlboro, Maryland urges the use of a collaborative approach for assessing programs in his paper, "A Practitioner's Perceptions Regarding Problems in Assessing the Effectiveness of Teacher Education Programs." This LEA staff development practitioner describes the multiple factors influencing the context of the new teacher entering the profession. The identification of appropriate competencies to be considered in the complex context of the beginning teacher is a very difficult endeavor. The implications of the needed competencies for staff development, inservice, or continuing

teacher education "offer a range of troublesome obstacles." However, characteristics of evaluation models which one should look for in this complex dilemma are suggested. Erly concludes with a plea for interinstitutional collaboration that would address the teacher preparation program and school district needs, with suggestions for how the collaboration might be supported.

"NCATE Accreditation: Problems, Issues and Needed Research," is written by William E. Gardner, Dean, University of Minnesota, and Chairman of the NCATE Council. Because Gardner was requested to focus on the role of NCATE in evaluation, the process of national accreditation in teacher education and the problems and issues in evaluating and accrediting teacher education programs are examined by him. He cites the serious difficulties inherent in the accreditation process. Thus, Gardner calls for attention to specific evaluation issues and concerns and for attention to the development of better techniques for doing evaluation. He also points to the need for research, including that which would examine the relationships between teacher education program criteria and teacher performance. The importance of developing a cooperative effort that would undertake analysis of various aspects of NCATE accreditation is sounded as a closing note to the paper.

Gary R. Galluzzo, Glassboro State College, New Jersey, answers the questions "What is the role of evaluation?" and "What is the purpose of evaluation?" to indicate the parameters within which program evaluation is conducted. Thus, in his paper, "Program Evaluation in Teacher Education: From Admissions to Follow-Up," he first addresses the role and purpose of program evaluation in teacher education. Galluzzo, in describing an evaluation model useful to him, includes the components of the model, the information deemed important to be obtained for each component and suggestions about how to collect the information. The four explicated components are admission, process, product, follow-up. The paper concludes with considerations for implementing the evaluation model in a teacher education program.

The Evaluators

Gary D. Borich, The University of Texas, Austin, addresses a vital concern in "Building Program Ownership: A Collaborative Approach to Defining and Evaluating the Teacher Training Program." The Borich paper approaches the evaluation of programs from an assessment of the need for a "meaningful sense of ownership in the teacher training program among all interested parties." Borich maintains that the problems experienced by institutions in evaluating their programs are not due so much to a lack of models but to the lack of a vehicle for engendering a feeling of program ownership. How to get faculty in divergent fields with divergent interests to care about more than their own limited piece of the training pie is the problem.

Thus, he presents an approach for a) thinking in a structured way about a teacher training program, b) working as a team with division and coordination of effort and roles, c) communicating a program definition in clear and concise notation which can aid in program design, development, and evaluation. The importance of the process of developing a program definition builds upon documenting the ways that teacher educators think about their programs and about the development and evaluation tasks that they perform, thus, aiding teacher educators both in their thinking and in their ability to convey their understanding to others. This understanding, in addition to considering and understanding the program's architecture or conceptual structure, makes it possible to bring about commitment, pride and ownership. In essence, communication with constituent bodies, the understanding of the nature and structure of programs and a thorough knowledge of the process of program definition seem to be essential ingredients for the coordination, integration and synthesis necessary for successful teacher preparation programs.

For this reason, Borich defines the work of program improvement as a social process and not as technology, feeling that the process of program design, development and evaluation has too long been ignored as the human interactive process that it is. The need for a better understanding of the conceptual structure of programs requires disciplined human interaction. No amount of measurement technology will

suffice. Borich provides a guide for the process of the human interaction.

In his paper, "Introductory Guidelines for Designing Evaluations of Teacher Education Programs," evaluator Nick L. Smith, Northwest Regional Educational Laboratory, Portland, Oregon, presents ideas for assisting those individuals who have had little or no experience in designing evaluations of teacher preparation programs. The guidelines would assist these persons in conceptualizing the process to a degree sufficient for engaging in meaningful discussions with qualified evaluation practitioners. The paper presents options to be considered in addressing five questions viewed as basic to the design of evaluation: What is to be evaluated? Why is the evaluation being conducted? What questions should be asked? What methods should be used? How good is the proposed design?

Smith maintains that each evaluation must be tailored to the local context in which it is to occur, the needs and interests of relevant parties and the resources available. The document is designed to facilitate the initial discussion and plans of individuals as they engage in the creation of effective evaluation design. The paper focuses on the evaluation design process only; other considerations, such as implementation and management of the evaluation, are reserved to space and treatment at a later time. Each of the five major questions comprises a section in the report, each section addressing what is to be decided, why the decision is important, what the possible resolutions to the decision are, and how a satisfactory decision can be reached.

Three key questions considering the evaluation of teacher preparation programs are addressed in the paper, "Exploration in the Evaluation of Teacher Education," by Daniel Stufflebeam, Western Michigan University, Kalamazoo, Michigan. Stufflebeam discusses three issues deemed important from his perspective:

1) What fundamental conceptualizations should guide attempts to evaluate teacher education?
2) What models are appropriate and useful for applying the recommended conceptualizations?
3) How should given models be operationalized?

Each question is addressed extensively in the paper; three additional questions about evaluation are reserved for future explication.

The paper comes from the position that the most important purpose of evaluation is not to prove, but to improve and that programs cannot work better unless their weaknesses are identified. For these and other reasons, teacher educators must subject their work to competent evaluation to sort out the good from the bad. In order to help teacher preparation practitioners do this, the paper has:

proposed a detailed definition of evaluation,

extended the definition to show how evaluation can be used to guide improvement efforts and serve accountability needs,

defined in detail the core concepts of context, input, process and product evaluation and how they relate to teacher preparation evaluation, and

provided general guidelines for designing evaluation studies.

Though Stufflebeam suggests that it may be easier for professors to give up academic freedom than willingly submit their programs to evaluation, he perceives that the teacher preparation field has readiness to foster quality, efficiency and public credibility through evaluation and that evaluation is a fundamentally important commitment to improvement and ways to make that happen.

A Response

Marlene Strathe is Director of the Educational Research and Development Center, University of Northern Iowa. Strathe was invited to write a reaction or "So What?" regarding the set of eight papers. The concluding paper provides her reflections. She extracted six themes which she perceived as woven through the papers. First, she identifies and comments on those themes and their implication for teacher education programs and their evaluation. Second, as a result of the six notions and her own concerns as a teacher educator practitioner, she suggests a number of directions which the agenda of program evaluation might follow.

This volume is the result of an intensive, interactive effort to look at program evaluation and its needs. The volume has been produced in order to share the current assessments of practitioner teacher educators in regard to evaluating teacher preparation programs, and also to present the most current, up to date, experience and research-based procedures which have been developed by evaluation experts to respond to concerns and problems of teacher preparation program evaluation users.

The papers reflect intense concerns, abiding beliefs and a consistent theme of urgency, to do something for the profession of teacher education that is practical, profitable, productive and will lead to improvement. While the authors express their ideas and concerns in strong statements of powerful persuasion, their points frequently are conveyed in an informal, familiar and conversational style. They tell their stories very well.

We hope this volume will be a useful resource to our colleagues who are undertaking the difficult tasks of program evaluation.

Shirley Hord
R&D Center for Teacher Education
The University of Texas at Austin

Tom Savage
College of Education
Texas A&M University
College Station, Texas

Lowell Bethel
College of Education
The University of Texas at Austin

September, 1982

TEACHER EDUCATION'S EVALUATION OF GRADUATES: WHERE ARE WE GOING AND HOW DO WE KNOW WHEN WE GET THERE?

J. T. Sandefur

Western Kentucky University

TEACHER EDUCATION'S EVALUATION OF GRADUATES: WHERE ARE WE GOING AND HOW DO WE KNOW WHEN WE GET THERE?*

J. T. Sandefur
Western Kentucky University

Introduction

In 1970, I chaired one of the first NCATE teams to use the new 1970 version of the Standards. Immediately thereafter I was invited to appear before the NCATE Council meeting in Biloxi, Mississippi, to describe the experience. I well remember advising the Council that I believed the Standards would fly, with the possible exception of Standard 5.1 and Standard 5.2, standards on evaluation of graduates and the feedback of those evaluative data into program development. I indicated that my team was presented data from a survey sent to graduates, who reported that their preparation program was the best that they had ever had, and surveys of employers, who suggested that the graduates of this institution were the best that they had ever employed. I further suggested that the techniques of evaluation in teacher education were seriously underdeveloped, and, in the absence of scientific and proven models for the evaluation of graduates, I predicted that future teams would be inundated with similar surveys of graduates who would readily admit that their preparation program was the best they had ever had, and cooperative local school administrators who would testify that these students were the best they could employ.

How I Became an Expert on Evaluation

I almost built a career on that one prediction. AACTE, recognizing my exceptional insight--almost approaching clairvoyance--placed me on

---

*The writer acknowledges the assistance of Ronald Adams in the conceptualization of this paper and of James Craig for assistance in the development of the research bibliography. Both criticized the paper more than the writer believed necessary.

their Commission on Standards. My first assignment was to attend NCATE Council meetings to monitor the standards causing trouble. Invariably I reported Standards 5.1 and 5.2, evaluation of graduates and use of evaluative data as the standards causing the most difficulty. My data were so impressive that I constructed a vertical bar graph that made Standard 5.1, in comparison to other standards, analogous to the Empire State Building in a row of ranch houses.

The Commission on Standards, recognizing the scientific accuracy of my data, concluded that there was only one viable solution to the dilemma of Standard 5.1. We would develop a series of descriptive case studies or alternate strategies to focus on the application of this standard. In fact, we used the term "illustrative models." As the newest member of the Commission, I was assigned the development of the first illustrative model which, incidentally, came from the printers as An "Illustrated" Model for the Evaluation of Teacher Education Graduates. I had several calls from readers wanting to know where the illustrations were and, if they had been inadvertently left out, would I send them.

My model, although unillustrated, was the only model to reach print. No one else came through. The fact that AACTE quickly sold out two printings was definitely not a tribute to the quality of the work. It was, rather, an indication of the desperate need of the institutions facing accreditation for help in coping with Standard 5.1.

That Was Prologue, What About Now?

The events I have described occurred in 1970-1972. Now it is 1982. What has happened to evaluation in the intervening years? What is the status and state of the art of evaluation of graduates of teacher education programs and what has been the impact, if any, of my "illustrated" model for the evaluation of teacher education graduates?

In response to my question, "What has happened in evaluation in the intervening years," the most significant thing I can think of is that NCATE Standard 5.1, Evaluation of Graduates, is now Standard 6.1, Evaluation of Graduates, and Standard 5.2, Use of Evaluation Results to Improve Programs, is now Standard 6.2, Evaluation of Results to Improve

Basic Programs. The practice of evaluation of graduates has made little advance in the past ten years. I make that statement as a practitioner who has more than a passing interest in evaluation. I make that statement as a member of the National Council for Accreditation of Teacher Education, despite the fact that Bill Gardner, the Chairman of NCATE, is in this audience and may hold a differing point of view, and I make that statement as an NCATE team chair for each of the past three years. The practice of evaluation is not significantly different in 1982 than it was in 1972. Institutions still rely primarily on surveys of graduates and employers. Furthermore, large research institutions are hardly more sophisticated than small teaching institutions. I surveyed the 33 institutional cases presented at the most recent NCATE meeting in Nashville to determine the standard most frequently cited as a weakness--I can tell you that you are already ahead of me--it was, of course, Standard 6.1, Evaluation of Graduates.

There is an area in which we have made tremendous advances in evaluation in the past ten years. I refer to the theoretical knowledge base about evaluation. I want to hold my comments about the significant advances we have made in this area until after I discuss the impact of my Illustrated Model for the Evaluation of Teacher Education Graduates.

I'm Really Not Negative, But...

I think the model made a tremendous impact--on Western Kentucky University. Two additional institutions adopted all or part of the model. Several institutions read the model. I believe it safe to suggest that they read it since AACTE sold out all copies, but the model made little or no impact on the practice of evaluation of students in institutions across the country. Let me cite reasons because I believe the reasons may also apply to models we will discuss during this conference and models that may be developed as a result of this conference:

1. The model was too complex and too unwieldy. I was pleased in reading Gary Borich's paper in which he described three models for evaluation: (1) The Needs Assessment Model, (2) The Relative Gain Model, and (3) The Process-Product Model, that I had included all except the relative gain component, and that only because I didn't know how. I

didn't believe that evaluation was capable of accomplishing that in 1972. Furthermore, I'm not convinced that it can be accomplished in 1982, except perhaps by researchers in laboratory controlled experiments. I eagerly await the development of the science to the point that all 600 NCATE accredited institutions can employ some validated technique for assessing pupil gain in relation to teacher competency.

2. The model was too expensive. In the nine years we conducted the study at Western Kentucky University, I estimate that we invested between a quarter and a third of a million dollars in evaluation--dollars that we are no longer able to get.

3. Although the model successfully evaluated graduates, albeit a complex, unwieldy, and expensive process, we were never truly successful in feeding the evaluative data back into the development of our ongoing teacher education program, at least not to the extent we had hoped. Although it must have been our fault, our faculty seemingly never understood the relationship between the data we collected and the ongoing revision of the teacher education program. Perhaps the fault lay in their lack of feeling of involvement and ownership. They seemed to think of the project as the Dean's and Ronald Adams' source of research and publications. Had we the opportunity to conduct the project over, we would have made provision for faculty involvement both in the decision-making and the conduct of the project.

On Second Thought

I would not leave this topic having you believe that our longitudinal evaluation study was not successful. It was. Under Ronald Adams excellent direction, the project produced dozens of publications and presentations. To my regret, it was better known nationally, and perhaps internationally, than it was known at Western Kentucky University. We not only learned something about the evaluation of graduates, we learned a great deal about how not to evaluate graduates.

Like the Farmer, We Know How to Do Better Than We Do

I said earlier that there was an area in which evaluation of graduates has made tremendous advances. That area is, of course, the theory and knowledge base relevant to evaluation. We know "how" to

evaluate far better than we evaluate. The past few years have advanced the knowledge base far beyond that which could be found in the literature in 1972.

We know, for example, that evaluation should be conceptualized as a process rather than a product. The idea that evaluation is an "instrument" or a "data set" or even a "method" is no longer appropriate--if indeed it ever was. Nor should evaluation be equated with research (Sommer, 1977); the two are very different processes. And while the methods and techniques of evaluation are an important component of the evaluation process in that they provide part of its foundation, they do not define it. In short, "good" evaluation methods provide a necessary but not sufficient condition for the conduct of program evaluations.

There are many possible methods and techniques that may be used when conducting program evaluations. Under the leadership of individuals such as Nick Smith (1981a, 1981b), meta-evaluation (i.e., the evaluation of evaluation) has been initiated on a broad scale and meaningful dialogue regarding alternatives (such as occurring here) has begun (Hord and Adams, 1981). The evaluation practitioner has many alternatives from which to choose in building an evaluation system for a particular program. The quasi-experimental and experimental approaches detailed by Campbell and Stanley (1964), the CIPP (i.e., Content, Input, Process, and Product) decision-making model of evaluation developed by Stufflebeam (1971), the procedures for conducting goal-free evaluations proposed by Scriven (1975, 1976), and the procedures for developing and using reliable and valid measurement instruments for "field" use developed by Brouch and Gomez (1977) and others represent some of the "tried and true" methods that are appropriate for use in many evaluation contexts. In addition, new, viable alternatives have been proposed that give evaluation practitioners new tools to consider. For instance:

1. Borich (1980) has proposed a needs assessment model for conducting program evaluations;

2. Patton (1980) and others (e.g., Fairbairn, 1982) have examined the characteristics and operational procedures associated with qualitative evaluation techniques;

3. Birnberg and Birnberg (1977) have outlined the zero-base budgeting process as a viable program review/evaluation

procedure to be used by non-profit organizations;

4. Guba (1981) has suggested investigative reporting as a model after which to fashion the program evaluation process;

5. Rossi, Freeman, and Wright (1979) have identified impact assessment techniques and the procedures for their use in conducting program evaluations;

6. Eisner (1981) has proposed that art criticism can be used as a model by which to review and refine the evaluation process before it is initiated;

7. Harper and Babigian (1971) have advanced provisional analysis as a means by which to review and refine the evaluation process before it is initiated;

8. Chafin (1981) has outlined the procedures by which ethnographic techniques can be used in program evaluation studies;

9. Todd, Kibler, Dodl, Walter, and Rollin (1975) have developed the Florida Assessment/Diffusion System (FADS) for the initiation, verification, formation, operation, and evaluation of educational innovations; and

10. Wolf (1975) has demonstrated that a legal, adversary model can be successful instituted as a model of evaluation.

These and many other alternatives are being developed, tested, refined, and incorporated into the evaluation systems--a much needed process. Which of the many alternatives is "best" must be assessed in terms of the particular situation in which the evaluation is being conducted; "best" is determined in relation to the evaluation situation, assuming, of course, that the methods and techniques possess three important characteristics: they are reliable, they are valid, and they are usable.

If a method or procedure is not reliable, valid, and usable, then from a practitioner's perspective a method will have no utility. Each set of data produced by a method or procedure must be judged according to the quality of the data and its usability in program decision making. We should not fall prey to Kaplan's (1964) law of the instrument (i.e., evaluate only those things for which our evaluation methods are suited) and, thereby, allow evaluation methodology to determine program goals, objectives, operational procedures, and so on. Quite the contrary,

evaluation methodology should be and must remain subservient to program concerns.

As can be seen from the sources quoted, the body of research on evaluation is expansive, recent; most of it within the past four years, and conducted by a new generation of researchers and evaluation experts. Thus my contention that the knowledge base more nearly approaches adequacy than ever before. Nevertheless, the practitioner is not putting this wealth of knowledge together into a practical, functional plan to effect the evaluation of graduates of teacher education programs.

So--What Can We Do to Build an Operational Evaluation System?

From our nine years of experience at Western Kentucky University operating a rather extensive evaluation system, we would suggest that there are only two major components of an operational evaluation system: (1) what we evaluate and (2) how we evaluate it. In case you think the problem has been understated, let me hasten to add that there are several subcomponents that must be taken into consideration.

I. What do we evaluate?

   A. Stated goals and objectives of the teacher education program provide the direction for the evaluation system. Not only must programs have clearly stated objectives, but these goals and objectives must be subject to evaluation and they must be made public.

   B. Sources of knowledge and beliefs

      1. Program goals and objectives should be based on organized knowledge and/or beliefs about effective teaching and/or effective teachers.

      2. Organized knowledge and beliefs upon which programmatic goals and objectives are based may come from several sources including:

         a. Research on teaching effectiveness
         b. Formal theories of educational practice
         c. Community, state, and national values
         d. Unsubstantiated, commonly-held beliefs about teaching effectiveness
         e. Logical, common sense education practice and "lessons" learned from experience

C. The context in which education occurs is an issue in evaluation of teacher education programs, particularly product evaluation, where one attempts to control for significant influences of contextual variables such as grade level, ability level, sex, socio-economic factors, administrative climate, etc.

II. How do we evaluate?

There are also considerations in "how" we evaluate:

A. The concept of evaluation must be viewed from a process perspective rather than from a product perspective. The idea that evaluation is an "instrument" or even a "data set" is no longer appropriate if indeed it has ever been.

B. Approaches to evaluation may differ, however, the relationship of the approach taken and the ultimate utilization of the data in program decision-making must be taken into account.

C. Measurement technique and instrumentation should be considered in the light of the approach to evaluation. Regardless of the approach, qualitative, quantitative, or whatever, each set of measurements must be judged according to the quality of data and its usability in program decision-making.

During our nine years of evaluation, we learned some things about the characteristics of evaluation. Things that experienced evaluators already knew. We learned, for example:

A. Evaluation of teacher education programs should be viewed as part of the teacher education program and not as a separate, external function assigned to another unit.

B. Evaluation should be systematic rather than random, purposeful rather than casual.

C. Evaluation should be a continuous process rather than one shot affairs.

D. Data gathering points should be planned to provide information at critical periods during preservice and inservice teacher development (Entry, Matriculation, Exit, Induction and Follow-Up).

E. Data of a longitudinal nature is preferred to cross sectional data.

F. Analyses and reporting of data should be planned in advance and modified with use.

G. Feedback and interpretation of data should be systematic and routine and viewed in light of program objectives.

H. Decisions regarding program change should be made with benefit of the evaluation data.

I. Implementation of the decision should be monitored and viewed as part of the program evaluation process.

J. The evaluation operational system should be reviewed periodically as to the relevance and usability of the data in decision-making and in relation to stated goals and objectives.

And, in Conclusion, What is My Message?

As far as evaluation is concerned, my message is "the 30-second shot clock has run 25 seconds, we have only about five seconds to get the shot off." Why would I use such an inept analogy? Because I believe it to be true. The states are rapidly pre-empting higher education's options to develop and implement effective evaluation systems. For example:

* Thirty-three states have taken some action relative to competency assessment of teachers.

* Twenty of these states have introduced legislation to accomplish the assessment. Of those states, twelve have passed the legislation, legislation is still pending on two, and eight failed to pass the legislation.

* In twelve states the department of education or the state board of education mandated, or directed in some way, the competency assessment of teachers. Action is pending in one state and under discussion in five others.

* Twelve states have required testing for admission and seventeen have testing required for certification. Eight states require testing both for admission and certification.

* Eighteen states require testing in basic skills areas, usually English and mathematics. Twelve require testing in professional skills and ten require testing in academic areas.

* Six states require performance assessment on the job, usually a probationary period of one or more years.

Why have the states taken this action? Because they believe we have failed to assure the public of the quality of our product. They believe that we have failed to exercise proper controls on entry to our

preparation programs, and they are assured that we have failed to establish qualitative standards for exit, therefore, they have assumed the responsibility to determine that only the able teach.

My state of Kentucky is one of those preparing to mandate competency assessment for certification and a one-year internship with on-the-job assessment prior to the receipt of continuing certification. We are presently moving toward an instrument identifying a number of competencies a teacher must have and a five-point scale on which each of the competencies will be rated. The rating will be conducted by the employing school district--a principal, a supervisor, and a resource teacher will make the judgment. A simple system? Indeed it is. An effective system? One can only guess at this time.

My point is obvious. Our failure to develop effective, functional systems to evaluate our graduates and to provide developmental feedback to our teacher education programs has resulted in an action that will place that responsibility at the state level. Of course, we must continue our efforts, primarily for program improvement, but our efforts will be secondary to the state's system of evaluation for certification purposes.

Let's Hear It For This Conference!

This conference deserves a 21-gun salute. The need for evaluation of teacher education graduates can in no way be overstated. But please, don't get tied up in complex models. Please don't speak in statistics. Please be aware that we already know how to evaluate better than we do. It is my fervent hope that our evaluators/designers can, in some exciting way, design an evaluation program that is simple, functional, and inexpensive. If this can be done, perhaps we can even help the states and regain our status.

Oh yes, I almost forgot to answer the questions I posed in the title of this paper, "Where are we going and how will we know when we get there?" Through the help of experts and knowledgeable people such as gathered here, we are going to design valid and reliable evaluation systems to apply to our graduates. We'll know when we get there because

the teacher education institutions will be using our systems, something that they have not yet done.

References

Birnberg, J. G. & Birnberg, S. C. The zero-base budgeting opportunity. Professional Psychology, 1977, 8(4), 442-450.

Borich, G. D. A needs assessment model for conducting follow-up studies. Journal of Teacher Education, 1980, 31(3), 39-42.

Brouch, R. F. & Gomez, H. Sensitivity, bias, and theory in impact evaluations. Professional Psychology, 1977, 8(4), 411-434.

Campbell, D. T. & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1963.

Chafin, C. K. Toward the development of an ethnographic model of program evaluation. Phi Delta Kappa/CEDR Quarterly, 1981, 14(1), 3-6.

Eisner, E. W. On the difference between scientific and artistic approaches to qualitative research. Educational Researcher, 1981, 10(4), 5-9.

Fairbairn, K. Fairness in qualitative evaluation. Research on Evaluation Program Newsletter, 1982, 4(3), 2-5.

Guba, E. Investigative journalism. In N. Smith (Ed.), New techniques for evaluation: Volume 2. Beverly Hills, CA: Sage, 1981.

Harper, D. & Babigian, H. Evaluation research: The consequences of program evaluation. Mental Hygiene, 1971, 55(2), 151-156.

Hord, S. M. & Adams, R. D. (Eds.). Teacher education program evaluation, 1981: Theory and practice. Austin, TX: Research and Development Center for Teacher Education, The University of Texas, 1981.

Kaplan, A. The conduct of inquiry: Methodology for behavioral science. San Francisco: Chandler, 1974.

Rossi, P. H., Freeman, H. E., & Wright, S. R. Evaluation: A systematic approach. Beverly Hills, CA: Sage, 1979.

Scriven, M. Bias control systems in evaluation. Paper presented at the annual meeting of the American Educational Research Association, 1976.

Scriven, M. Evaluation bias and its control. Occasional Paper 4. Kalamazoo, MI: The Evaluation Center, Western Michigan University, 1975.

Smith, N. L. (Ed.). New perspectives in evaluation - Volume 1: Metaphors for evaluation. Beverly Hills, CA: Sage, 1981.

Smith, N. L. (Ed.). New perspectives in evaluation - Volume 2: New techniques for evaluation. Beverly Hills, CA: Sage, 1981.

Sommer, R. No, not research, I said evaluation! APA Monitor, 1977, 8(4), 1, 11.

Stufflebeam, D. L. (Ed.). Educational evaluation and decision-making. Itasca, IL: Peacock, 1971.

Todd, W. R., Kibler, R. J., Dodl, N., Walter, D., & Rollin, S. A review and critical analysis of the literature pertaining to the diffusion of educational innovations. Paper presented at the annual meeting of the International Communication Association, Chicago, IL, 1975.

Wolf, R. L. Trial by jury: A new evaluation model. I. The process. Phi Delta Kappan, 1975, 57, 185-187.

REQUIREMENTS OF A DATA BASE FOR EFFECTIVE PROGRAM EVALUATION

Robert A. Roth

Michigan State Department of Education

REQUIREMENTS OF A DATA BASE FOR EFFECTIVE PROGRAM EVALUATION

Robert A. Roth
Michigan State Department of Education

In order to discuss the requirements for a data base for program evaluation, there are several key areas which warrant consideration. I would like to comment on several of these, particularly in the context of program approval processes as conducted by state agencies. Although I will focus to some extent on the state context, you will probably note that for the most part the issues relate to all contexts and are thus not necessarily unique to the state perspective. I also will give considerable attention to practical concerns, as opposed to conceptual issues, whenever possible in order to facilitate practical outcomes of our model building.

Why We Need It

The first area of interest I would like to address is why we need to evaluate. There are various purposes of program evaluation, and several different groups who have an interest in evaluation. These, in turn, have an effect on what information is collected and how decisions are made. An institution of higher education (IHE), which is a school, college, or department of education (SCDE) in this instance, conducts program evaluation for a variety of reasons. The most important reason should be in order to improve programs and the quality of its graduates. Let us not forget this purpose because it is germane to our entire purpose here, and I shall come back to it.

SCDE's also may evaluate in order to differentiate among parallel programs, particularly when a choice has to be made among these. We also find evaluation being conducted because of external pressures, be they political (from the legislature), fiscal (from funding agencies), required for existence (state approval) or desired for prestige (NCATE). SCDE's interests thus range from survival to status.

The public and consumers of our product certainly have an interest in evaluation of programs. Teacher education students and parents have a vital interest in the quality of educational personnel we prepare. Interestingly enough, however, this type of information is seldom asked for by consumers, and judgments about program quality are not made on the basis of data but rather on subjective considerations. The contributions of personnel preparation are thus not separated from other variables, such as conditions in the schools, and thus an accurate assessment of preparation quality is not made.

I need not point out to you the tremendous criticism teacher education has been subjected to in recent years from all quarters. The professional literature and popular press are full of harsh assessments of teacher education. A former college president wrote an article for a major city newspaper and titled it "Schools of Education Must Stop Being a Joke" (Bunzel, 1981). The Kappan contained an article "Why Teachers Can't Teach" (Lyons, 1980), which charged that "teacher education is a massive fraud."

Incidentally, the article received the prestigious National Magazine Award for public service. This reflects how severe the criticism has been. This not only demonstrates a critical need to evaluate programs but it also presents an opportunity. If rigorous evaluation is conducted, credibility may improve as well as programs. When programs can demonstrate positive outcomes, this provides the information and opportunity to respond to critics.

State government also has a very strong interest in program evaluation. The state protects the public interest. With mandatory schooling, the state has a responsibility to insure that the schools have educational personnel who are appropriately prepared to be in classrooms. Since certification is generally based on program approval systems, evaluation of programs is the basis for the entire state educational personnel system. Without a strong program evaluation procedure, teacher certification is virtually meaningless.

State agencies, SCDE's, teachers, the public, and the teaching profession as a whole all share an interest in certification reciprocity. An effective system of reciprocity requires quality assurances. The Interstate Agreement on Qualification of Educational

Personnel is a true reciprocity system based on interstate contracts involving states. Those contracts specifically require that each state conduct evaluation of institutions and their programs.

These various agencies have an influence on the type of evaluation which is conducted. Formative evaluation is called for when our interest is primarily program improvement. Summative evaluation is needed by funding agencies and when selecting among alternative programs. We thus must ask ourselves very early in the design process why we are evaluating, for what purpose.

I would also like to mention why we need a data base. It is obvious to me that decision making of any type is best made on the basis of data, and program evaluation is essentially a decision making process. The key question then becomes what the nature of those data should be. I will address this question very directly later.

Purposes and Problems With Accreditation/Approval

If we look more closely at the area of accreditation and program approval, we find that it too has varied purposes. Some of these are as follows (Roth, 1979-80):

1. Determine if the institution meets a "minimum acceptable level" as defined by program evaluation criteria.

2. Determine if institutions meet quality standards of form. These standards are at a higher level then "minimum," and quality is operationally defined in terms of form (or inputs and processes) such as faculty degrees, rather than substance (outcomes).

3. Determine if institutions meet quality standards of substance, such as the quality of instruction (regardless of degrees held by faculty) and the quality of graduates (regardless of the number and type of courses required).

4. Verify whether or not an institution has made appropriate progress toward achieving its own goals.

5. Provide for the improvement of programs and stimulate higher quality.

6. Encourage programs to lead society through innovative viewpoints of schooling and teacher education, as opposed to conforming to traditional modes of teaching and teacher education.

7. Provide for prescriptive evaluation through precise standards.

8. Provide for descriptive evaluation with more broadly stated standards assuming there is no one best way.

9. Merely validate information submitted by institutions.

In addition to examining purposes of program evaluation and accreditation, we should also review what we now do and some of the problems we are encountering. Again I would like to address this in terms of accreditation and state program approval. Some of these problems are as follows:

1. Clarity of Purpose. Lack of clarity results in confusion over why it is being done; the roles of those involved; and expectations of outcomes of the process. These factors promote resistance. With clarity of intent standards can be consistent with purpose, and procedures can lead to the desired outcome.

2. Appropriateness of Criteria. These indicators of program quality form the basis for the entire evaluation. If the criteria are faulty, no matter how effective our processes or evaluations are, we will not have a meaningful evaluation.

3. Cost. Costs of all types of program evaluation have been a major problem, be they perceived or real. Program evaluation is frequently seen as an expense with little return.

4. Impact. Of what value is accreditation and state approval? Questions have repeatedly been raised as to whether or not the process makes a substantive contribution to the improvement of programs. Outcomes in terms of the cost, or cost-effectiveness, is a key consideration.

5. Legal Issues. Education is currently experiencing a significant increase in the number of court challenges and discussions of issues related to legal authority and constraints, both in K-12 schooling and in teacher education. Accreditation and certification are not exceptions and will increasingly have to face challenges related to the issues and questions arising from a legal context. One area of concern is the nature of evidence used in accrediting or approving programs. Freeman (1976, p. 104) indicates that there are four issues related to evidence. They are: (1) kinds of evidence; (2) reliability of evidence; (3) relevance of evidence; and (4) sufficiency of evidence. Each of these need to be explored from a legal perspective in order to determine the ramifications for future court actions.

6. Diversity of Institutions. Teacher education institutions vary considerably across a number of characteristics. The various ways in which they differ are according to size of institution, resources available to institutions, the student body composition of institutions, whether or not they are privately or publicly funded, the level of highest degree that is offered, the type of campus (main or regional), and type of mission, such as teaching, research or service. A problem which then results is how does one allow for such institutional differences in a common accreditation process. Does one use the same criteria for all institutions? If one compares the programs at one institution to others, which institutions will form the basic reference group for the comparison?

   The long standing criticism of the flexible ruler or "rubber standard" concept reflects a concern related to differences in institutions. The problem has materialized in the application of standards, where some of the larger more sophisticated programs are held to more stringent standards than are the smaller programs with limited resources. It should be noted that whether or not this is a problem will depend upon the basic purpose of the accreditation process. If it is to determine whether or not indeed the institution has achieved its objectives, then it may be appropriate to have multiple sets of standards for varying types of institutions. If one is concerned with all institutions meeting certain rigid standards, then inequities resulting from the application of standards to differing degrees for different institutions becomes more apparent.

7. Complexity of Process. One of the more frequently voiced concerns regarding accreditation is the complexity of accreditation as a process. A great deal of time and effort are involved in selecting an accreditation team, in organizing the team, and scheduling the visit. An even greater amount of time is spent at the institution developing the self-study report which requires volumes of data, some of which may never be used by the accreditation team. Additionally, the on-site evaluation process itself is somewhat complex, involving the use of standards, subjective judgments, and a variety of types of data. If one looks at the data and the process, it is necessary to determine whether or not indeed the data and judgments are valid. Do they measure what they are supposed to, are they truly representative of the programs (internal validity), and are the findings applicable to all graduates of the institution in a given period of time (external validity).

   The question of reliability is also one of interest. There is considerable inconsistent application of standards particularly if there is not some type of common training, or common interpretation of standards. In addition, there is a problem with the relationship between the standards and the process used to evaluate the standards. The problem

which we are faced with is to try to simplify this process without reducing the effectiveness of accreditation.

8. Evaluators. The individuals who are selected to serve on teams will vary considerably, depending upon their experience in teacher education and with evaluation techniques. Some of these individuals may not be qualified. Sometimes they are not necessarily selected on the basis of their qualifications as much as they are selected since they represent specific interest groups. Frequently, evaluators will use their personal experiences as the basis for comparisons and will substitute their subjective evaluations for objective data and objective criteria for assessing such data. The problem of qualified evaluators casts doubts over the reliability of the process.

9. Overlap and Coordination. Many institutions are subjected to at least three types of accreditation or program approvals. The first is a regional accreditation which deals with the overall quality of the institution. In addition, most states have a program approval process which includes a mandatory on-site evaluation by a team selected by the state. The third type of approval is that of voluntary national accreditation, which is conducted by NCATE.

   If one examines these procedures it is readily apparent that there is a great deal of similarity in both processes and standards. Hence, there is a great deal of overlap of effort and the same data are frequently reviewed by three different teams at three different times. This results in considerable costs to the institutions.

   The problem of coordination to eliminate unnecessary overlap is a cost-effectiveness question which needs to be reviewed.

10. Public Disclosure. Many individuals feel that once an accreditation or program review is conducted, particularly those conducted through public agencies, that such data should be made available to the public so that appropriate decisions can be made about institutions and their programs.

NCATE and Program Deficiencies

NCATE and its findings may provide insights into the needs for program evaluation. The NCATE standards cover a wide range of evaluation areas, having standards for inputs and processes (form) and outcomes (substance). The results of NCATE accreditation reviews also provide information on program weaknesses and weaknesses in institutional evaluation efforts.

One aspect of NCATE accreditation procedures which is quite revealing is the weaknesses cited in programs which have been denied accreditation. Since NCATE will be described in another paper, I will only briefly summarize the relevant findings to illustrate my point. The introduction to NCATE Standard 6 states that "The ultimate criterion for judging a teacher education program is whether it produces competent graduates to enter the profession and perform effectively" (NCATE, 1977, p. 10). This standard requires that the institution develop a well-defined plan for evaluating the teachers it prepares, and that it use the evaluation results in improving the program. In separate studies Sandefur (1971) and Fritschel (1975) did analyses of NCATE reports and Board reviews, concluding that there was little evidence of systematic evaluation of graduates, that the greatest concern of the NCATE Evaluation Board was with evaluation of graduates and use of these results, and the greater number of weaknesses cited were for the "evaluation" and "use of results" standards.

In NCATE Update (NCATE's newsletter) it was reported that "For the second consecutive year, evaluation of graduates and use of evaluation results occupied the top reasons for denial of accreditation of advanced programs" (NCATE, 1981, p. 6). For basic programs these standards 6.1 and 6.2 ranked first and third (NCATE, 1981, p. 7) among weaknesses cited for both those programs denied and accredited in 1980.

Common Purpose and Need

As we look at each of the areas discussed: purposes of evaluation and program approval; problems encountered with accreditation and state program approval; and deficiencies in institutions as determined by NCATE studies, it appears that there is a common area of intent which has also been the greatest area of weakness. This is the need to focus on outcomes of teacher education programs, that is, the quality of graduates.

Earlier I stated that the most important purpose of program evaluation, or reason why SCDE's would want to evaluate their programs, would be to improve the quality of their graduates by improving programs. NCATE standards state that the ultimate criterion for judging

a teacher education program is whether it produces competent graduates who enter the profession and perform effectively. This is consistent with the state responsibility of protecting the public interest, and the needs of the consumer who wants assurances of quality. Also, what better way to respond to critics and establish the credibility of teacher education than by providing evidence of competence of graduates, rather than by offering evidence of form. The purposes of evaluation and the needs of all the interest groups are best served by this approach of evaluating the product.

I quite frankly believe we cannot afford any other approach; otherwise teacher education will remain at the bottom of the academic heap and continue to be the subject of the most vile criticism. Very strong actions must be taken, and rigorous evaluation of our graduates is one such action.

To provide further rationale, consider the relationship between program evaluation criteria and the effectiveness of graduates. When we evaluate factors such as faculty, facilities, and administration, we assume that these contribute to our ultimate goal of preparing graduates who are effective in the classroom. There is quite a gap between these input variables and the resultant product. If we measure characteristics of graduates themselves, however, one need not make as great an inference about the relationship between what we evaluate and effective teachers. This is depicted on figure one.

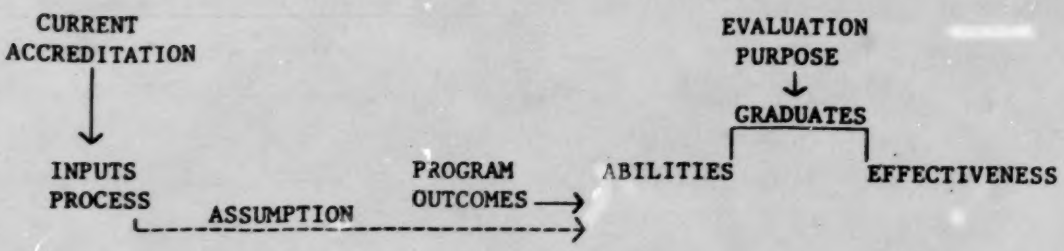


FIGURE ONE

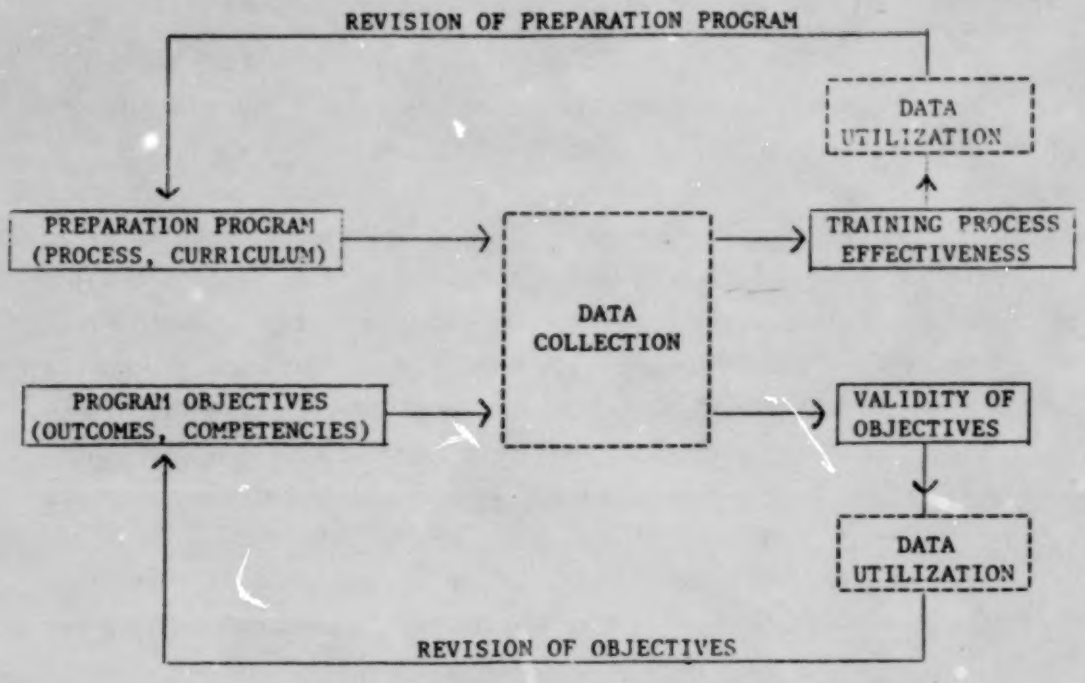
In spite of this logical, political, and practical need for evaluation of graduates, evidence suggests that this is the weakest part of program evaluation. If our purpose is to improve the quality of our graduates and our programs, then why do we not evaluate our graduates? There are at least three major reasons which immediately come to mind. The first of these is the problem of methodology. We need to develop procedures to collect data on effectiveness of graduates. Further, we need to make them accessible and easy to use, even for those procedures we now have. In other words, better dissemination of practical usable systems or packages. Recently, some such systems have emerged (Roth, 1980-81-82). Workshops funded through state and federal grants, and seminars by professional organizations are appropriate media.

A second reason for lack of evaluation of graduates is financial. Institutions are either not willing to make this commitment or they are unable to do so. We thus need to devise methods which are reasonable in cost, and an important part of this is efficiency of evaluation.

A third reason why we have insufficient evaluation of graduates is that quite simply we do not require it! Our state program approval procedures are notably weak in this regard, although there is some interest in it (Roth, 1979, ATE). NASDTEC standards, for example, are silent on this issue, and NCATE has not focused on this as a priority.

For years I have observed awards being given to "exemplary" teacher education programs. These have been based on a variety of process type criteria, such as school-college collaboration. What I would like to see is an award for teacher education programs based on evidence of effectiveness of the program in terms of competence of graduates, or on the basis of its system for collecting and using data on graduate effectiveness.

Earlier I said I would address the question of the data we need for decision making. There are two major areas we could look at, the preparation program itself (processes, curriculum), and the program objectives. Data collected on each of these areas provides different information. Figure two shows that data collected on the preparation program tell us something about the effectiveness of the training process. These data would be used to revise the preparation program.



PROGRAM EVALUATION CYCLE

FIGURE TWO

Data collected on program objectives tell us something about the validity of these objectives. We then can answer questions such as: Are the established objectives appropriate? Are some of these objectives irrelevant or of little importance? Are there other objectives which should be included. These data would be used to revise the objectives of the program.

When we talk about measuring outcomes of programs we should be looking at both knowledge and performance. The use of examinations for certification is increasing throughout the country. These data, however, could also be used for purposes of program approval. In addition, however, we need to have feedback on programs from those who have been prepared in them. This could be accomplished through follow-up studies of graduates. We should also have data based on observations of graduates. This could be collected through follow-up studies also. Using questionnaires to ask for opinions or ratings from supervisors or administrators is one possibility but has many short-

comings, e.g., inter-observer reliability, lack of definitive criteria, etc.

A well structured system for observation of graduates would contribute significantly to improving this type of information, and would improve the data base we would have for program approval decisions. A method for collecting reliable observation data is currently being devised by Coker, Medley, and Soar, and their initial work is encouraging.

I believe that we need to establish some outcomes oriented state program approval systems. I believe this data base should be derived from several sources and provide several types of outcome data.

It is also important that the system be designed to insure that such data, once obtained, can readily be used and indeed are used for program improvement. A model of how this might be incorporated into a state system is shown in figure three.

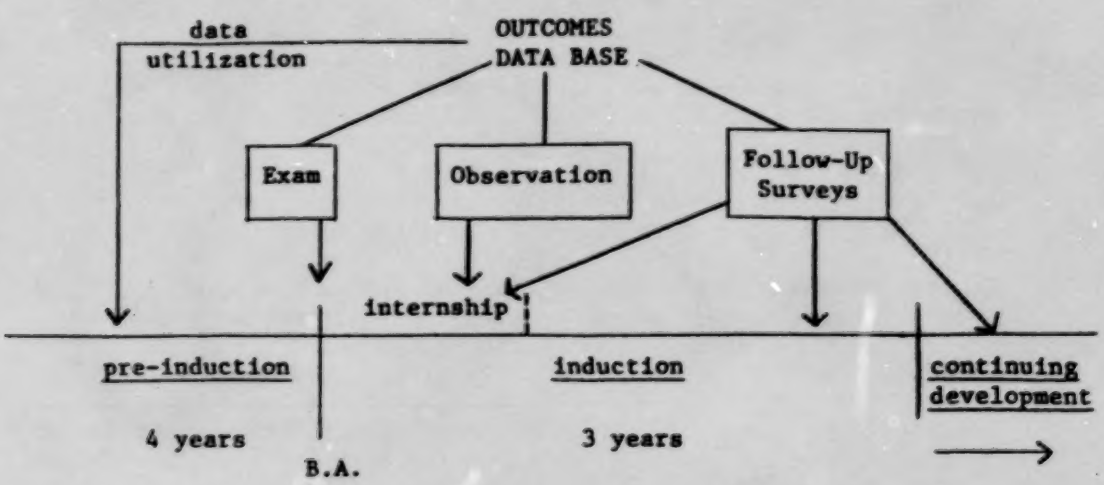


FIGURE THREE

One of the problems we encounter is the clarity of the descriptors of quality. Without clearly describing what it is we are evaluating, it is difficult to obtain appropriate evaluation data. SCDE's need to better define what it is we want to measure and evaluate. A statement we frequently hear, and I regretably cannot give credit to the originator of this specific statement, goes something like this:

"Many outcomes of teacher education programs cannot be measured or stated in measurable terms.

These are the most important outcomes.

These are the qualities in which our graduates excel."

Suggested Criteria for Evaluation Design

The following issues are suggested for the design of a program evaluation/approval model. These are based on the needs identified throughout this paper.

1. The model(s) should be designed with full recognition of the purpose of evaluation which it addresses. If varied purposes could be addressed this should be clearly identified. If the model(s) could be adjusted for varied purposes this would increase its usability.

2. Program evaluation model(s) should focus on outcomes of preparation programs. We need to further develop these.

3. Program evaluation model(s) should take into account a variety of types of programs, such as preservice and inservice preparation programs.

4. The model(s) should facilitate interpretation and use of data collected in order to improve the program. Techniques for this process should be incorporated into the model(s).

5. Utility of model(s) is important if they are to be widely used. Feasibility or ease of use of model(s) is important so that a variety of types of institutions with varying degrees of evaluation sophistication can readily utilize the model(s).

6. Model(s) may be developed at different levels of sophistication so that institutions might select an approach which best fits its level of expertise, available funds, etc.

7. It may be necessary to develop models which are less rigorous in design in order to accommodate institutional restrictions, to insure use by institutions with limited faculty and financial resources.

8. For program approval purposes, model(s) should be designed which may be used by institutions and monitored by the state.

9. Distinctions need to be made between model(s) to be used by institutions versus model(s) used by state systems.

10. Program evaluation model(s) should be cost-effective, so that impact of the evaluation is clear and justifiable in terms of cost. Institutions with large numbers of graduates and varied programs are particularly faced with this problem.

11. Program evaluation model(s) should identify appropriate times and intervals for data collection to insure collection of data when and only when it is most important and to avoid unnecessary data collcation.

12. Economy and efficiency of evaluation needs to be considered. The fewest number of variables which provide a reasonably accurate description of program quality should be sought out.

13. The model(s) should strive for development of criteria for each decision category (program elements).

14. Sampling of graduates is the least understood and given the least attention in program evaluation. In terms of methodology it is the greatest area of weakness and needs further development as an integral part of program evaluation models. Sampling is a particular problem for institutions with a large number of students in varied programs.

15. Methods of using data from tests of graduates' knowledge need to be developed.

16. For state systems of program approval, methods of sampling programs as well as students should be considered.

17. A means of incorporating and weighing various types of data in arriving at an overall conclusion should be part of the program approval system.

18. Evaluation model(s) should strive to diminish the variance of evaluator differences to enhance reliability and portability of models.

19. Methods of documenting the contribution and value of the evaluation model itself should be built into the model.

20. State program approval systems in particular need legally defensible models whose procedures, logical basis, and criteria can be defended.

References

Bunzel, J. H. Schools of education must stop being a joke. The Cleveland Plain Dealer, November 3, 1981.

Freeman, L. D. Character and use of evidence in program approval. Licensing and Accreditation in education: The law and the state interest. Lincoln, Nebraska: Study Commission on Undergraduate Education and the Education of Teachers, 1976, 104.

Fritschel, A. L. Program weaknesses identified by NCATE evaluation boards using the 1970 standards. Journal of Teacher Education, Fall, 1975, 26(3), 205-209.

Lyons, G. Why teachers can't teach. Phi Delta Kappan, 1980, 62(2), 108-112.

National Council for Accreditation of Teacher Education. Evaluation, program review and planning: Evaluation of graduates, Standards for the Accreditation of teacher education, NCATE, 1977, 10.

National Council for Accreditation of Teacher Education, NCATE Update, NCATE, Washington, D.C., February 15, 1981, 3(1), 6, 7.

Roth, R. A. Alternative approaches to state program approval. Action in Teacher Education, Certification/Accreditation in Teacher Education, The Journal of the Association of Teacher Educators, 1979, 1(3,4), 53.

Roth, R. A. The varied purposes of accreditation and program approval. MATE Viewpoints, The Journal of the Michigan Association of Teacher Educators, 1979-80, 2(1).

Roth, R. A. The program evaluation instructional series, Books I-VI, University Press, 1980, 1981, 1982.

Sandefur, J. T. An illustrated model for the evaluation of teacher education graduates. A report to the American Association of Colleges for Teacher Education's Committee on Standards, October, 1971.

A PRACTITIONER'S PERCEPTIONS REGARDING PROBLEMS IN ASSESSING THE EFFECTIVENESS OF TEACHER EDUCATION PROGRAMS

Maurice C. Erly

Prince George's County Public Schools
Maryland

A PRACTITIONER'S PERCEPTIONS REGARDING PROBLEMS IN ASSESSING THE EFFECTIVENESS OF TEACHER EDUCATION PROGRAMS

Maurice C. Erly
Prince George's County Public Schools
Maryland

The development of teacher education programs, from pre-service through in-service, offers a challenging set of obstacles. In this paper I will attempt to establish a context into which a newly graduated teacher enters the teaching profession, in which he continues his professional growth. We will then look at some problems associated with competency assessment and their implications for the methodological structures for performance evaluation. Some counsel and guidance will then be provided for constructing some evaluation models. Enveloped within these remarks will be some implied issues regarding my role as a local educational agency staff development professional.

Today teacher education and performance occurs within multiple context of change, accountability, and professional conditions which significantly impact the outcomes of teacher performance and program development. Within the context of change we are currently having to deal with changed attitudes, support systems, expectations and society's value for education. Attitude change is reflected primarily in the reluctance to accept the movement in the 70's towards innovation and innovative practices as well as a significant reduction in the degree of respect accorded a professional educator. Change in support is reflected primarily in the monies allocated for the acquisition of resource materials and personnel assistance for the conduct of training and performance in education. This reduction in support is generally couched in the need for holding fast the real estate tax of citizens. It is also reflected as a direct result of a reduction in pupil enrollment in the public school system. The change of expectation is significantly different than the expectations held in the 1960's and early 70's. These expectations are toward a more specific outcomes-related performance base and in a constricted set of expectations for

services to be delivered by schools to the individual child. Another significant change in expectations is reflected in the reduction in support for what had previously been perceived as the "humanistic" teacher who now oftentimes is labeled by the critics of education as the "whishy whashy" creative teacher. Change in society's value of education, in large part, is tied to the quantitative expression of pupil performance (therefore teacher performance), as expressed in the second context which I wish to address, which has been labeled accountability.

The measurement of teacher performance and student gains in the late 70's and early 80's has been expressed through the use of numbers rather than letters. The alphabet which was used to describe the performance of a youngster has been supplanted with the mathematical model of analysis which brings every performance criterion up through measurement through quantitative expression. This movement toward quantified achievement has significantly reduced emphasis in curriculum areas for which numbers cannot readily be assigned to student performance. The second consideration under the context of accountability is the management by objectives move which began in the mid-seventies. This context, much like the quantitative approach, requires a very specific response for every set of proposed activities. It is almost the parallel of numerical specificity through verbal specificity. Implicit in this move has been the need for pre-planning of an activity, with the pre-statement of what one intends to do, followed by the action of "doings," followed by an evaluation system which is asking how well you did what you said you were going to do. This set of activities requires very specific statements of purpose and achievement, with a parallel match between them. The third area within the accountability context is the whole notion of cost effectiveness with the decrease in the amount of monies and personnel to carry on the job of schooling. There has been an increased emphasis on performance at minimum cost. This context has permitted the reduction of fat from the skeleton, and often generating the question, "why aren't you doing as well as you used to?" The final accountability context is the increased emphasis on planning and reporting in the classroom, in the school building, in the school system and at the university. It is

almost as if we don't really believe what you say you are doing unless you can show that you had planned to do that, and the plan had been approved. This context has developed in some a significant feeling of distrust and a general feeling of not being of much value.

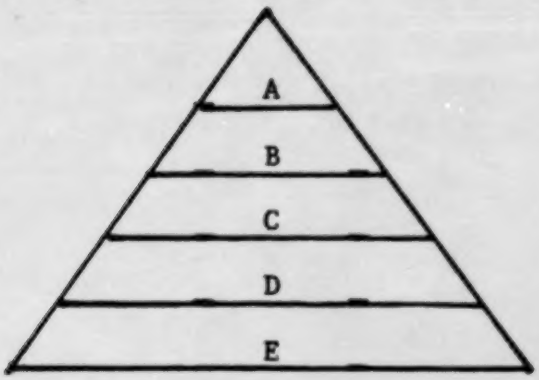
A third area in which we can look at the context of education is that of professional condition. There is no question that in many of the major school systems there are fewer material and direct assistance resources being made available in support of the teacher. In this age of increased emphasis on staff development there has been a creeping decrease in the amount of directed assistance provided to teachers. Another significant professional condition is that of lower overall educator's morale. This is perceived by some as being a direct result of the reduced support from the community for what is happening in education, the reduction in psychological strokes received by teachers who are attempting to do a good job, and the national trend of reducing the number of teachers employed by some of the major school systems. Given these forces we hear reported a higher number of professionals who are performing with a personal anxiety and professional dissatisfaction for the job they are being asked to perform. We see an increased number of shifting assignments requiring educators to relocate in other school facilities on a bi-yearly basis. For many who do not relocate we see them receiving increased numbers of assignments out-of-field of training. Teachers are expressing dissatisfaction with receiving more prescripted curriculum for implementation with less opportunity for creative decision making and expression. We see a work force which is older and less mobile being charged with carrying on the more demanding business of education.

Given the context described above the teacher educator today is faced with the demand of developing a more competent educator. This same teacher educator worked within many of the same societal contexts described above. In this age of change, accountability and professional condition, educators are being asked more directly to prove that they make a difference in developing significant learners. The teacher educator is being asked likewise. Thus, we are faced with the requirement of developing teacher education programs which are to be competency based and which can make some assessments regarding how

effective a teacher the institution's graduates will be. When you consider the multitude of teacher education programs for which competency based programs should be developed, an institution can easily be millions of dollars invested in a requirement which is indeed most complex. It seems that if one should address performance competency of an educator those competencies should be basic competencies appropriate across teaching disciplines and job assignments and indeed appropriate to many other vocational choices which may lie outside of the area of education. Rather than develop 1,010 competencies which a mechanical engineer should have (United States Naval Academy instructional program), it seems to me that the investment of intellect and energy should be focused on competencies which have been identified as basic or general to education as well as other vocational choices. These basic (core) competencies should be very specific in nature. They should be performance based, and of value to the professional as an individual as well as an educator. They may be such simple competencies as an ability to observe, to cite objective statements of what has been observed, synthesizing objectively observed incidents, recognizing value influences, or recognizing distortions of fact. They may be such competencies as being able to logically sequence incidents or events. The point being made is that all of the competencies listed have relevance in value in a wide-range of setting both inside of and outside of education. They cross discipline and level assignments within a school and are appropriate to one's personal life as well. Such simply stated competencies quickly become complex when one begins to examine how a competency is applied when it is performing a job function in an educational setting. Synthesizing the complex environment, context of activity, and the availability of data upon which one acts is a significant challenge to the evaluator attempting to assess the competencies of an individual. If the assessment system is to provide a measure of how well one can perform within a competency area it must be able to examine the operational levels at which one performs within the given competency. Listed below is a triangle which has been labeled "Stages of Competency Assessment." Stages A through E represent a movement from the practice of the competency in its simplest form through the practice and assessment of the competency in its more

complex form. It is a visual representation of the multiple levels at which one might be asked to perform within a competency area. Level A and the space within it demonstrates the need for the least amount of time, effort and resources for assessing the competency at that level. Each succeeding level requires more time, space, and planning for assessment. The mere measure of a competency at Stage A or B does not permit the validation of one's capacity to perform that competency in the myriad of conditions under which he is asked to perform.

Stages of Competency Assessment

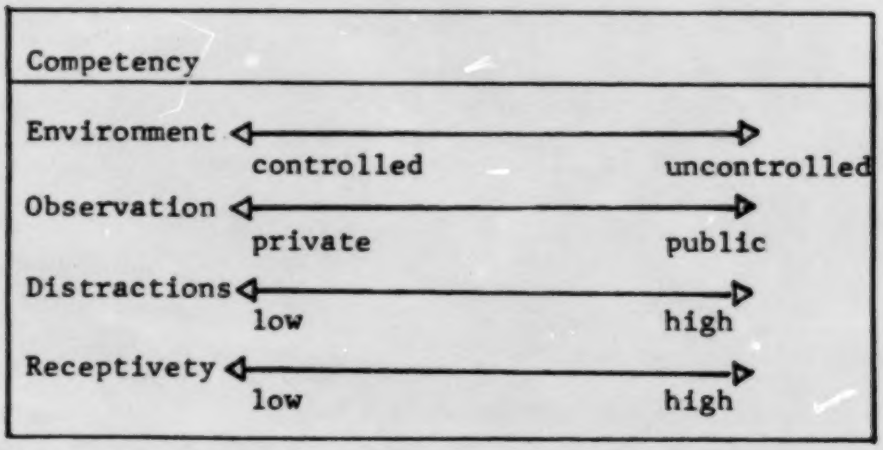


A. "Core" Competency
B. Reconfigured competency application in single context
C. Application of competency integrated within compound context
D. Application of competency integrated within complex context
E. Combination of competency within single and compound context

To further confuse this problem of assessment you should look simply at Stage A of the competency and apply to it some additional factors of complexity. These factors are not intended to be totally inclusive but are offered solely for the purpose of simply describing the problem of competency assessment. If we took the simple notion of objectively reporting an observation and applied the complexity factor of environment one can quickly see that the degree of which one is competent in objectively reporting what they have observed in a

controlled environment can be significantly different than how well one objectively reports what is observed in a very uncontrolled or chaotic environment. If we applied a second factor of complexity which oftentimes can occur simultaneously with other factors we can see that observing and then objectively reporting in a private or secluded setting can be quite different in its accuracy than one which is in a public setting which provides a wide range of additional stimuli which may not be under one's control. Lay on to this a third factor of complexity in which distractions are interfacing with the other factors, and one readily sees that the ability to perform that basic or core competency called objectively reporting what was observed is indeed most difficult. If you choose to enter even a fourth factor and that being a human factor called receptivity we can see that how well one is judged as having objectively reported what is observed is extremely mixed or distorted.

Factors of Complexity



What has been attempted to this point is take a simple core

competency and share how it becomes most complex when moving it through the stages in which you might experience it. The stages become more difficult when one begins to add the factors of complexity one faces when trying to perform that competency. The core competency which was used to illustrate the stages of competency assessment and the factors of complexity are genuinely replaced with perceived needs which are not as specifically stated. In Appendix A you will note that the perceived needs listed in the "Demographic Analysis of Perceived Needs Survey" have multiple meanings. What is implied when an educator indicates a need to improve in "dealing with disruptive students?" What are the range of knowledge of skills and competencies which one employs when attending to this problem or concern? How does one begin to teach each of these knowledge and skill competencies and how does one effectively assess the level at which the competency has been developed by the learner? Quickly we see that an area of concern which many teachers may face has a host of meanings and prescriptive approaches which the learner must acquire and effectively implement. Hopefully, this example begins to reaffirm the concerns I have for the whole notion of competency assessment if it is permitted to be played out as thousands of specific generic needs which are primarily focused to educational conditions and concerns. This perceived need of "dealing with disruptive students" could possibly be filtered down into some core competencies that are employed not only in dealing with this particular need but in other perceived needs listed on this demographic analysis such as:

Dealing with class size
Working with the mainstreamed youngsters in a regular classroom
Relating to large numbers of students on a more individualized basis
Assessing special needs of students
Counseling with students that have special needs
Using ability grouping within classrooms
Managing classroom organization effectively

The above listed perceived needs and concerns are an outgrowth of a staff development survey and study conducted in our school system this current year. These are some of the needs and concerns that were expressed by a cross-section of our professional staff. These were then worked into statements which collected many of the specific statements

which permitted this broader statement and which were administered through the staff development survey. The implication for developing effective staff development activities and their concommitant evaluation programs offer a range of troublesome obstacles.

Given the social, political and economic content in which training educators occurs and the complexity of the task of assessing to what degree one is able to perform, it is difficult to offer counsel and guidance for constructing evaluation models which ignore either of the two sets of obstacles listed. Characteristics of evaluation models which are not normally the focus of educational assessment but for which I think serious consideration should be given are as follows. Do the evaluation models permit an incremental reaffirmation of one's self-worth? Do they provide increments of improvements which are realistic within the social context of contemporary teaching? Do they focus on knowledge and skill competencies which are basic to the educational act, and transportable to future career choices outside of education? Are they developmental and cover a career stage of four to eight years beyond one's bachelor's degree? Are they appropriate to performance certification for career choices inside and outside of public education?

In examining the design of evaluation models, one should look at items such as follows:

Does the evaluation system demonstrate a time and resource efficiency appropriate to the investments permitted to educational systems for self-improvement?
Is the assessment approach considered of low cost?
Are the assessment results easily interpretable?
Can the results of the assessment be easily communicated to the public?
Are the assessment outcomes specific rather than general, and focused on knowledge or skill competencies?
Are the areas being measured recognized by both educators and the public as being valuable?
Does the assessment process require special services from outside of the organization for whom the evaluation was designed?

Within the context of education, cited earlier in this paper, it is my belief that these considerations are important. Current as well as future educators must be able to readily recognize that the evaluation systems, through which they are being asked to pass, are structured to

determine the levels of basic competencies of the educator. In this regard there are several areas of professional development which have required extensive inservice and staff development programs because the necessary skills and knowledge were not acquired during preservice training. These areas primarily deal with the Specific Learning Disabilities Program of our school system, which has taken into account the multiple learning styles of youngsters and the prescriptive teaching strategies necessary for effective instruction. A second area requiring extensive assistance is the area of classroom management of both materials and persons, as well as basic classroom organization and reorganization.

Teacher preparation programs and school district needs can be better addressed if there is an interinstitutional collaborative effort between the concerned parties. This interest should be demonstrated through the development, implementation and evaluation of programs not only at the IHE but also at the LEA as well. This interinstitutional collaboration should occur between the LEA and its primary IHE. Such a relationship should permit the extensive utilization of school sites for internships, practicums, field observations and student teaching. This is best accomplished if the commitment for such a relationship is affirmed through the placement of key personnel of both institutions on a regularly convened task force.

APPENDIX A

DEMOGRAPHIC ANALYSIS OF PERCEIVED NEEDS SURVEY
in Rank Order of
Professionals with 1-2 Years of Experience*

Perceive Need	Years of Experience 1-2 N=110	3-8 N=875	9-15 N=1800	Above 15 N=1830
Dealing with disruptive students	41.0	41.1	44.8	38.1
Working with parents on behavior problems	30.9	26.4	26.8	19.3
Motivating students	27.3	29.8	32.4	24.8
Dealing with class size	26.7	39.7	41.4	32.1
Finding the time to grade papers	23.2	39.3	39.5	31.5
Working with mainstreamed students in the regular classroom	21.4	26.6	31.0	25.7
Being aware of available resources to program effectively for students	21.2	12.9	12.0	9.1
Relating to large numbers of students on a more individualized basis	21.0	35.4	37.0	27.9
Adapting materials for students with special needs	20.5	23.8	23.5	18.5
Assessing special needs of students	19.5	20.6	18.4	14.0
Being aware of various discipline techniques	18.6	16.6	15.1	11.2
Preparing classroom instructional materials	18.4	22.4	22.6	17.6
Evaluating the effectiveness of classroom teaching strategies	17.6	10.5	11.3	9.1
Counseling with students who have special needs	17.4	16.6	19.1	14.4
Using ability grouping within classrooms	14.9	16.9	17.6	12.6
Managing classroom organization effectively	14.9	15.2	15.1	10.6
Developing skills to conference successfully with parents	13.7	10.6	9.8	7.6

*This chart is based upon respondees who selected responses 5, 6, & 7 indicating experiencing a medium to major problem with the listed concerns.

NCATE ACCREDITATION:

PROBLEMS, ISSUES, AND NEEDED RESEARCH

William E. Gardner

University of Minnesota

52-B

NCATE ACCREDITATION:

PROBLEMS, ISSUES, AND NEEDED RESEARCH

William E. Gardner
University of Minnesota

This paper will examine the many issues and problems in evaluating and accrediting programs in teacher education. It will suggest that there are several obvious and serious difficulties which beset the process of accreditation in this field. These problems demand that attention be given to the development of better evaluation techniques and the conduct of additional research on a number of factors, including the relationship between criteria for teacher education programs and teacher performance. I will begin by discussing the process of national accreditation in teacher education, then move to a brief analysis of the problems and issues I perceive in this process, and finally, point to the specific evaluation needs occasioned by this situation.

Introduction

For about 30 years, the National Council for Accreditation of Teacher Education (NCATE) has accredited programs in teacher education in the United States. While other agencies (most notably the American Psychological Association and the American Library Association) have also claimed and won the right to accredit school-connected programs in their specialities, the generic accreditation offered under the NCATE banner has prevailed and the Council recognized by the regulatory agencies of the Federal government (The Council on Postsecondary Accreditation and its precursors) as the responsible agency in teacher education.

In the beginning, NCATE was produced by a coalition of enlightened leaders representing the American Association of Colleges for Teacher Education (AACTE), the National Education Association (NEA), and the chief state school officers. These agencies began to act collectively to promote quality standards and attempt to protect the integrity of teacher education through a process of "peer" assessment and self-

regulation. The influence of NEA was not especially strong at the outset, but has grown in recent years to where it equals or surpasses that of AACTE.

While the number of accredited institutions has increased steadily, a minority of schools, colleges, or departments of education (SCDEs) belong to NCATE. Only 40% or so of the total number of teacher preparing institutions have volunteered for NCATE accreditation. Included in the group accredited by NCATE are a majority of the larger institutions; probably as many as 80% of the people entering teaching complete programs approved by NCATE.

In several critical ways NCATE is dissimilar from accrediting agencies in the more prestigious professions like law and medicine. In virtually all states a prospective practitioner cannot sit for the state examination without first completing a program at an accredited law or medical school. In education that simply is not the case, a fact which greatly diminishes the significance of NCATE actions. Further, in other professions control of the accrediting process is firmly in the hands of practitioners. While law school professors or medical school deans, for example, may be part of the accreditation scene, it is clear that control is in the hands of the "bar" or the medical examiners. Only in the most exaggerated definition of the terms "control" or "practitioner" could NCATE be considered under the control of the profession. Rather, governance of NCATE bears some resemblance to a parliamentary government with several equal-sized factions; in NCATE's "parliament" there are three factions (AACTE, NEA, other constituent organizations, e.g., NCTM, NASP, etc.). The passage of policy or the making of accreditation decisions depends upon some coalition between and among the members of the three groups.

NCATE differs in some degree from other accrediting agencies also in terms of its basic purpose. While accrediting agencies are both "developmental" in the sense that they attempt to help an institution improve its programs, and "regulatory" in that they try to drive out inferior programs, accrediting agencies in many other fields tip the scale toward the regulatory end. They do have the power to control entry into the profession because graduates of nonaccredited places cannot practice. NCATE has operated until quite recently under what

could be called a policy of "watchful stimulation;" the Council sought to aid the cause of teacher education by pointing out weaknesses in programs by allowing provisional accreditation to be awarded pending changes in the program. This policy has changed rather dramatically in the last few years. The Council is now far more aggressive. It has declared itself to be the "standard bearer" for the consumer, abolished provisional accreditation, toughened its evaluative criteria and increased substantially the percentage of negative accreditation decisions. It should be borne in mind that NCATE is not effectively linked with the process of certification and licensing in states and, hence, is a weak accreditation process compared to law and medicine. Nonetheless, these recent changes have been dramatic and not without serious consequences for IHEs. The fact that NCATE is basically "toothless" does not remove serious embarrassment and/or public relations problems for institutions denied accreditation.

The recent surge of self-consciousness and muscle flexing by NCATE was accompanied (and perhaps caused) by a nation-wide increase in the institutional approval activities of state governments. The states, of course, have always possessed the legitimate power reserved to them to regulate higher and, hence, teacher education. But until very recently, the policies of state departments toward teacher education could best be described as one of benign (in some instances, malevolent) neglect. Within the past decade, however, most of the 50 states have established review processes to evaluate both institutions and programs in teacher education. Whether conducted through the state bureaucracy (the department of public instruction or state departments of education) or through teacher-dominated boards of teaching, these efforts bear an amazing similarity to what NCATE does in that they employ similar evaluative processes and standards. It is true that state approval processes are mandatory and exist solely for purposes of consumer protection, while NCATE is voluntary and, as noted before, has both developmental and regulatory goals. Nonetheless, the NCATE process and those of the various state agencies constitute the primary external tension on programs in teacher education.

It is the basic contention of this paper that both processes, state and national, are necessary and overall beneficial to teacher education

and its clients, but that they contain serious flaws which threaten their integrity and, hence, the overall concept of quality control in teacher education. It would be legitimate to analyze both NCATE and state processes, but for purposes of clarity and simplification, I will concentrate almost entirely on the NCATE process. In all likelihood, the problems and issues identified here are as true for state approval processes as for NCATE.

NCATE's Process of Evaluation

The process employed by NCATE involves a maximum of five steps.

1. An institution decides to apply for accreditation, and prepares an institutional report (IR) which covers each program presented for accreditation. The IR explains and describes the program in some detail, indicates how the institution attempts to meet all of the NCATE standards, and provides a set of base-line data about the program and the institution.

2. The NCATE office appoints a team chair who leads an evaluation team to campus for no more than three days. During the visit the team conducts interviews, examines records and other documents, interviews students and faculty in an attempt to "validate" claims made by the institution in the IR. Before the team leaves campus, it determines by consensus whether the institution meets or does not meet each of the NCATE standards and whether the institution has strengths or weaknesses in any of them.

3. Under the leadership of the team chair, the Visiting Team prepares a written report containing the results of the Team's consensus on the standards and its analysis of the quality of the programs at the institution. The institution receives a copy of this report and may file a rejoinder presenting evidence that the team may have overlooked or challenging interpretations or criticisms made by the team.

4. At two of its three regular annual meetings, the Council takes up accreditation cases and decides whether each institution's programs are accreditable or not. Audit committees composed of three Council members are formed to review materials from the several team visits and prepare recommendations to the Council based upon their audit.

Institutions are quickly informed as to the action taken by the Council in regard to their programs.

5. Institutions which are denied accreditation may have Council decisions reviewed by an appeals board. The appeals board may rule in favor of the institution or it may reaffirm the Council's decision. In either case, the appeals board makes a recommendation to the Council, which then takes final action.

The Standards

The process just described is intended to assess whether an institution has met NCATE standards. These standards are of two types, one at each degree level. In NCATE terminology, for example, "Basic Standards" are those applied to entry level programs while "Advanced Standards" are applied to graduate or post-baccalaureate programs. The two sets follow similar formats and have similar content, although there are some minor differences reflecting the needs which apply to one or another of the levels specifically.

Each set of standards is organized in six broad categories, reflecting a conventional curriculum development approach. Institutions are to determine objectives, decide who the students will be or describe who they are, collect appropriate faculty and physical resources to support the program, and evaluate the results in a systematic fashion. In addition, the entire enterprise must be under the control of a "designated unit" of the faculty. Basically, then, the standards are concerned with the governance of teacher education, its curriculum, the way an institution selects and utilizes faculty and students, the physical resources devoted to teacher education, and its evaluation.

The process for revising standards or adopting new ones is relatively easy. The Council revises or adopts by a 2/3 vote at a regular meeting, assuming that the proposed change has been "disclosed" for a four month period. This ease of revising encourages frequent changes in standards. According to Alan Tom (1981), there has been a marked increase in detail and number of standards in the past two decades. In 1960, for example, the statement of standards contained about 4,000 words, while the current ones are more than 50% longer and

the 1970 standards mid-way between the other two. At the present time, there are more than 25 standards in each set compared with 22 in 1970 and only 7 in 1960.

With only a few notable exceptions, the standards are "process" rather than "product" oriented. That is, they mandate that institutions have a given process in place instead of demanding that students acquire competencies in certain prescribed areas. The standards dealing with admission, retention and advising of students, for example, ask only that a procedure exist for each of these functions rather than describing the desired characteristics of students at entry or exit from the program. Likewise, the curriculum standards mandate that programs be designed to elicit certain behaviors linked to an institution's conception of the teaching role; it does not demand that a given set of objectives be adopted for the program.

How are these standards validated? As noted above, standards are adopted by a 2/3 majority vote of the Council. Since the voting membership of the Council is made up of its three largest constituents in equal proportion, it is clear that a new standard must be acceptable to more than one of the constituent groups. The NEA or the AACTE alone simply does not have the votes to pass a new or revised standard. All standards, then, could be said to possess a kind of "political" validity in that they must be produced through compromise, lobbying, or horse trading. Several standards have an obvious "construct validity." One of the canons of good educational practice, for example, is that all faculty have appropriate educational credentials and be assigned to work in their areas of expertise; few would quarrel with this construct or principle as a criterion for a teacher education program. Other standards lack construct validity in which case they probably represent a kind of consensus feeling among members of the Council that a given topic or issue (e.g., multicultural education or education of the handicapped) or that a given procedure (e.g., including students in decision making) must be present in good educational programs for future teachers.

As a group, NCATE standards are most accurately described as broad process goals which outline some desirable directions for teacher education. Clearly, they are not criteria containing operational

definitions. On the whole, the standards have questionable validity, at least in the classical meaning of that term and, hence, there is a serious reason to be concerned about the reliability of the judgments made on the standards.

The Visiting Teams

These standards are applied to the teacher education programs of institutions by the Visiting Teams described earlier. In the last analysis, the Visiting Team Report (VTR) is the major determinant of the Council's accreditation action. As noted, the Council does have other data on which to base its decision (e.g., the IR, the institutional rejoinder to the VTR), and at times, the Council overturns the findings of the team. But exceptions here merely prove the rule; a well-written, convincing VTR is very persuasive with the Council. They key team role obviously is that of the chair. The chair sets the tone for the team, plays a major role in assigning and directing the evaluation activities of team members, and has the major responsibility for writing the VTR and for being the liaison between NCATE and the institution for all accreditation matters.

In any typical year, NCATE teams will range in size from about 5 to 16, (with a mode of 9), depending on the size and complexity of the programs being examined. Because teams are appointed with regard to the need for approximately equal representation from the various constituencies in the NCATE family, assembling an NCATE team is an exercise demanding solomonic wisdom and considerable conceptual dexterity. A typical team will contain representatives of AACTE, NEA, and the specialty constituent organizations in roughly equal proportions and will include as well women, minorities, and students. Typically a team will contain no one who, before the visit begins, knows anyone else on the team, and it would be extremely unlikely that anyone on the team would do another team visit during the same year.

The NCATE schedule most often calls for the team to arrive on a campus on Sunday evening, do its validation and evaluation work on Monday, Tuesday and Wednesday morning, and depart Wednesday afternoon following an exit visit at which team judgments are made known to the

institution. Deviations from this schedule are rarely made, even when an institution has a full range of graduate and undergraduate programs up for accreditation. Accommodation to the size, complexity, or special needs of a given institution is made by expanding the size of the team and/or designing a person as an assistant chair.

Problems with the NCATE Process

All systems of evaluation have problems and, of course, NCATE is no exception. One of its basic problems flows from the fact that teacher education in the United States is a huge enterprise. There are some 1350 SCDEs with state approved programs in teacher education. These include some of the most prestigious, the richest, the largest, the weakest, the poorest, and the smallest in the land. Although a minority of the nation's SCDEs are accredited by NCATE, the magnitude of the task dwarfs that of any other accreditation agency, not to mention the resource base of NCATE. Of the 530 or so institutions accredited by NCATE, each is on a seven year cycle; thus, about 75-80 institutions must be visited by NCATE teams each year. This number will be increased to some degree by the fact that some institutions typically request accreditation for new programs or for program additions on an "off cycle" basis. Since NCATE meets only twice a year to consider accreditation decisions, at each of those meetings the Council on the average must act on the applications of 35 to 40 institutions during the two day meeting. The raw material for the Council's actions is supplied by different teams for each of the institutions, teams which include about 700 individuals drawn from diverse populations--deans and professors (primarily education but including the liberal arts as well), teachers, administrators, and students. Such accreditation teams, melded from disparate groups, contain people who are professionals in the respective fields, but who are basically unskilled, undoubtedly untrained, part-time volunteers.

The sheer size and complexity of the problem, then, raise questions about the validity of assessments made for NCATE. Can the Council's criteria and standards have similar meanings and be applied equally and fairly across all of the institutions visited?

Such a question could more easily be answered affirmatively if the standards were uniformly clear and unambiguous. Such does not appear to be the case. Each standard does ask institutions to indicate how a given task or function is performed. For example, the practicum standard of NCATE (2.3.4) requires that the institution provide its students with "direct, substantial, quality participation in teaching over an extended period of time." While this seems clear, this standard, like all the others, also contains requirements that the program demonstrate somewhat more. As Wheeler notes, "On one level the basic requirements are straightforward and general: a task or function must be performed. But in each component of every family of standards there are requirements that call for the program to demonstrate a high level of performance" (Wheeler, 1980, p. 22).

Wheeler claims the difference between evaluations at the two levels is really the difference between, in his terms, the "presence and absence approach as opposed to the indepth approach." The former places heavy emphasis on whether a task or a function contained in the standard has been performed at all while the latter is concerned fundamentally with how well it has been performed. His observation and analysis indicate that some Teams use only the "presence and absence" while others probe more deeply into the qualitative dimensions of the program.

The point is that the standards contain a host of interrelationships and subtleties. Teams likely interpret these standards differently and, hence, provide differential evaluations. This situation constitutes the insidious "rubber ruler" so graphically described by deans of institutions denied accreditation.

Another problem with the NCATE standards involves what one of my colleagues calls, "the implementation of wishes and dreams." This is the deplorable tendency to write standards which embody noble goals but make demands far beyond the limit of our knowledge about teaching or teacher education.

A proposed rule governing bilingual teacher education in Minnesota, for example, enjoins IHEs to "apply teaching methods to different ways of learning taking into consideration how differences in culture affect learning." Such a requirement ignores the fact that there is only a very limited descriptive literature pointing to the interaction of

culture and learning and that literature is not sufficiently developed to provide a basis for making the kind of diagnostic, clinical assessments called for in the rule. Several NCATE standards share this characteristic. The standards on multicultural and special education (among others) call for institutions to develop competencies in their students which are not well-specified and/or which exceed the existing knowledge base. Without taking issue with the intent of these rules or their status as laudable goals for American schools, I must note that our knowledge base is simply not powerful enough to enable us to work the kinds of miracles these standards require. Except in the most general sense, we do not have the knowledge and skill needed to make prescriptive, clinical judgments.

Improving NCATE's Process of Evaluation

This paper has been sharply critical of NCATE, its standards, and process of evaluation. I must stress that NCATE accreditation has many strengths and that the organization does much good work (See, i.e., Comments by Wheeler, 1980). Even the standards which seem most confusing and ambiguous (e.g., the governance standard, the multicultural education standard) have provided nonetheless a national yardstick for teacher education programs and, as such, they have provided a measure of protection for teacher education programs which otherwise would be at the mercy of parochial state interests.

In his presidential address to the AERA in 1980, Michael Scriven made a devastating attack upon virtually all types of evaluation, reserving some choice words for that carried on under the accreditation banner. Among the various flaws and fallacies he noted were:

> . . . a disastrous failure to look at the right criterion variables (they always look at process instead of outcome variables); essentially, total failure to look at interpanel reliability; no effort to avoid halo effects, no standardization of language or procedures, no calibration training or testing of participants, no systematic study of the indirect costs or side-effects of programs, no follow-up, no critical review of the process by clients or their representatives--or no visibility for such reviews, and so on. (Scriven, 1980b, p. 8)

Does Scriven's criticism hold true for NCATE? It's tempting to answer with a crisp 'yes,' but in truth we simply do not know. The recent study by Wheeler concludes that NCATE evaluation does in fact discover many of the worst problems and denies accreditation to some of the marginal teacher education programs. In addition, Wheeler found the recent NCATE efforts at self-improvement and stricter standards to be impressive. Of course, he also uncovered inconsistencies in the application of standards and inconstancy in the behavior of some evaluators (Wheeler, 1980).

I agree with Wheeler's judgment that certainly some good things happen through NCATE evaluation, but recognizing serious weaknesses in the process, I conclude with Scriven that this "...aspect of professional activity could stand much more serious analysis than it has ever received" (Scriven, 1980, p. 8).

Beyond making the obvious suggestions of telling NCATE to improve its standards, or better train its teams, what can the evaluation profession say to this accreditation process? Specifically, what kind of serious analysis is needed? Are there models which can be adapted to NCATE's needs? What research should be undertaken on the NCATE process?

Rather than try to answer all these questions or make comprehensive recommendations to meet NCATE needs, I will list four issues which need analysis and which describe the areas of dialogue needed between those concerned with the NCATE process and evaluation specialists.

1. One of the most critical issues is reflected in Scriven's outrage at the lack of product assessment in all forms of accreditation and program approval. His deep concern is that the performance of graduates escapes review; it should, he claims, be the main focus and not just the "ghost at the banquet" (Scriven, 1980a, p. 112). The clear implication of this point of view is that performance, after all, is the central purpose of professional training and all other considerations should be subordinated to its measurement. And accreditation decisions should be made with regard to the success graduates experience in "doing" their profession.

One can hardly quarrel with the contention that performance is of central importance, but one can raise serious doubts about our ability to define all aspects of teaching performance and then measure that

performance as a function of a training program. At the risk of oversimplifying, I think we might agree within fairly broad limits on some aspects of what constitutes the technical dimension of teaching, and agree as well on how to measure the ability to perform technical functions (lesson planning, control of verbal behavior, reinforcement techniques, and so on). But there are other teacher behaviors than technical ones and other metaphors for teachers than "teacher as technician." As Alan Tom notes, this metaphor denotes a person skilled in producing student learning, but it ignores the normative aspects of teaching -- the teacher's sensitivity to his/her role in moral development, in making curricular decisions about what knowledge should be taught, and in understanding the dynamics of school and classroom (Tom, 1981). In other words, teaching has a "gestalt" which involves more than the technical aspects of the role.

There is little, if any, consensus on these normative matters, on what constitutes appropriate teacher behavior and how to measure same. Until some agreement is reached as to what constitutes the knowledge base, it's difficult to see how we improve evaluation very much by assessing only the more narrow and technical outcomes of a program. Rather desperately, and for a variety of reasons beyond the needs of accreditation, we need research identifying those aspects of teacher behavior which can be related to training programs.

2. Another consideration for accreditors and evaluation experts is what to do about the number and complexity of the NCATE standards. As noted before, there are many standards, they exist at two levels, many are particularistic, and they are not without ambiguity. An examination of the standards would certainly lead to two conclusions: NCATE's approach to improving its standards has been to add more of them; and it is extremely unlikely that any two NCATE teams could measure any standard in exactly the same way or provide the same operational definition to a standard. At the very least, we need concerted expert attention devoted to an analysis of the standards to determine whether operational definitions can be given to any of them or whether it is possible to develop a smaller number of more powerful standards.

Evaluation experts might contend that such analysis should be done by those responsible for adopting the standards and, so, the Council

itself should undertake the effort. Unfortunately, that's impossible. The political system of NCATE has only functioned to increase the number of standards, never to cause their diminishment or even their close examination. Moreover, any suggestion by higher education to revise by deletion is viewed suspiciously as self-interest. What we need is the involvement of evaluators in the development and testing of different sets of standards for teacher education.

3. Another area of consideration is the question of who should serve on NCATE teams and how they might or should be trained. I am convinced on the basis of my observation of and participation on about ten such teams that the people who serve are bright, without guile or bias, and anxious to do a good job. Unfortunately, they (we) are also amateurs and therein lies the rub. Clearly there is need for the development of training modules or packages which go beyond the anecdotal level of current NCATE training packages, and we desperately need help from the evaluation profession.

Beyond this, however, we should also examine the question of whether NCATE teams should be composed of "evaluation specialists" rather than "discipline specialists." The problem may not be a training problem (how to get the discipline specialists to do it better) but a conceptual one (perhaps the wrong people are doing the assessing). Should this matter be a question of preference only? What does the literature say about the relative merits of evaluation by the two kinds of specialists? What research should be done? What models constructed?

4. There is increasing concern over the costs of the total NCATE process. Although the NCATE visit itself is relatively inexpensive, the overall cost to the institution is quite high. The self-study should involve substantial faculty time and the production and printing costs of the report are of increasing concern. Moreover, the costs of visits are only part of NCATE's expenses, and institutions and the constituents of NCATE must contribute money to support the organization's infrastructure. As the ravages of inflation and budget cuts take their toll, the financial condition can only worsen.

To date, attacks on this problem have recommended only that minor adjustments be made in the current structure of NCATE--smaller teams, longer periods of accreditation, fewer meetings of the Council, and the

like. It would seem time to consider some radical alternatives to the whole process.

What might these be? Again, the literature is not particularly rich on this topic, although it does contain provocative examples. One such is Henry Levin's examination of the cost utility of three methods of determining teacher performance (accreditation of programs, testing, and observation). His study, an exercise in the economics of information, suggested that the cost/utility ratio of testing was considerably lower (e.g., more efficient) than the other two.

Levin's work is not at all conclusive, but it does suggest that there are other evaluative techniques and procedures which should be examined. Again we need the consideration of this issue by evaluators and education researchers.

In summary, then, I think we need a cooperative effort to analyze the degree to which NCATE can/should evaluate product as opposed to process, the ambiguity of its standards, the composition and (perhaps) the training of its team, and the relative costs of accreditation against other methods of reaching the same goal. This is a very tall order, but a very important one.

References

Levin, H. Teacher certification and the economics of information. Educational Evaluation and Policy Analysis, 1980, 2(4), 5-18.

Scriven, M. The logic of evaluation. Inverness, CA: Edge Press, 1980 (a).

Scriven, M. Self-referent research. Educational Researcher, April, 1980, 7-11. (b)

Tom, A. An alternative set of NCATE standards. Journal of Teacher Education, 1981, 22(6), 48-52.

Wheeler, C. NCATE: Does it matter? The Institute for Research on Teaching, Michigan State University, Research Series, 92, November, 1980.

PROGRAM EVALUATION IN TEACHER EDUCATION: FROM ADMISSIONS THROUGH FOLLOW-UP

Gary R. Galluzzo

Glassboro State College

67

PROGRAM EVALUATION IN TEACHER EDUCATION:

FROM ADMISSIONS THROUGH FOLLOW-UP*

Gary R. Galluzzo
Glassboro State College

The purpose of this paper is to outline the needs of one teacher educator as they relate to program evaluation. But before any discussion can begin, and the types of information which should be included in an evaluation are delineated, a few basic questions must first be answered. Answers to the questions, "What is the role of evaluation?" and "What is the purpose of evaluation?" set the outermost limits, or provide the parameters, within which program evaluation will be conducted.

The role of program evaluation in teacher education should be to create and stimulate critical discussion about the goals and objectives of a teacher education program. The clients of the evaluation are the faculty, the students, and the administrators who participate in the program. The discussion generated by the evaluation will provide common working definitions of key principles, thus curtailing un-focused debates on less than clear terms, thereby clearing the way for meaningful dialogue. What is meant by "cognitive outcome" or "performance assessment," as well as how each is measured, should be clear to all. In short then, the role of program evaluation is to create an environment wherein evaluation generates discussion and is recognized and accepted as an integral part of the teacher education program.

The second question, "What is the purpose of program evaluation in teacher education?" has a two-part answer. The typical answer is "because we want to make decisions about the effectiveness of the program." It is accepted that if we know the strengths and weaknesses of the program, we are in a better position to make informed decisions about improvement. This purpose can be labelled the "program improve-

---

*The author wishes to thank Dr. Ted White for his thoughtful assistance in the preparation of this paper.

ment purpose." Most institutions build their evaluation plans around the program improvement purpose and therefore focus on various selected outcome indicators, e.g., GPA, SAT scores, and others. It is assumed then that program evaluation should be formative evaluation which seeks to improve the process of becoming a teacher, and to increase the knowledge, skills, and attitudes of preservice teachers. However, there is a second, and more basic, purpose to program evaluation. This basic purpose is concerned with processes or transactions of the teacher education program. What is implicit in this dual purpose approach is that program evaluation that is limited to follow-up studies, such as typical interpretation of NCATE's Standard 6, or limited to product/outcome-only assessments, fails to provide a thorough understanding of the teacher education program, and how entering students acquire those desired teacher behaviors. This is not to indict performance assessment or follow-up surveys, for they are essential and must be included in any program evaluation effort. However, there is more to evaluating a teacher education program than product or outcome only. By collecting data on the processes or transactions of the program, the decision-makers are in possession of much more information which will assist in their improvement effort. The second purpose of evaluation then, is to gain a description and understanding of the processes within the teacher education program. All too often, programs are changed without knowledge of the effectiveness of the part to be replaced. The data called for in this paper create a portrayal of the process of becoming a teacher and its relation to the observed outcomes. The next part of this paper will present an outline of the evaluation, what information is important, and how it might be collected.

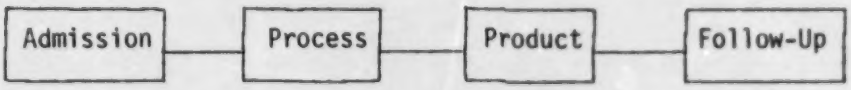
Overview

It was indicated above that a model for program evaluation should be a comprehensive model that starts its data collection when students are accepted into the professional education program. It should continue throughout the duration of the teacher education program and include process as well as product measures. The evaluation model

should terminate with a three-year follow-up effort in which graduates are selected at random for mailed survey or classroom observation. Thus, the model in its skeleton form is at its maximum a seven-year model for institutions which accept students as freshmen, and a five-year model for those institutions which accept students as juniors. The model has four parts (see Figure 1) which includes admissions, process, product, and follow-up. Within each of these four areas, the question, "What is important to know?" must be addressed. This question is answered below for each component of this skeleton model.

Figure 1

Skeleton of Program Evaluation Scope



Admissions

For the most part, teacher education programs have specific entrance criteria. These typically relate to the college's admissions standards and may include SAT or ACT scores, rank in high school graduating class, and in some instances, scores on a state test of basic skills, usually reading, writing, and math skills. Some institutions include a measure of a student's reasons for choosing a major in education or any other attitude or personality measures. It would seem valuable to have this information, not for admissions only but as descriptive data about the characteristics of each incoming group. McNergney and Carrier (1981) describe a handful of measures which teacher educators can employ in a longitudinal study, such as locus of control, conceptual level, and teaching anxiety. That is, a selected attitude or personality measure could be administered two or three times throughout the evaluation cycle to monitor the changes in the students' scores over time. While data about admissions are valuable to understanding any program, the beginning of a student's program presents an excellent opportunity to begin longitudinal studies of the college experience and the teacher education experience. The purpose of these descriptive data is to provide a picture of who selects a major in

education. Particularly helpful would be information about the students' attitudes toward teaching children and other personality variables, such as, authoritarianism or dogmatism. Placed on a longitudinal design, these measures help to reveal the shifts over time from entry through student-teaching and possibly follow-up. Then correlations may be run between performance and personality characteristics.

Process Evaluation

The next area of the model is the process phase of the evaluation. Once the student has been accepted into the program, the process phase begins. There is rich data that is typically left out of program evaluation. It is important to know what the program is like from the students' viewpoints and the faculty members' viewpoints. For example, it would be helpful to know their opinions and expectations about the relationship between their campus courses and their field experiences. Questions such as, "What are the field experiences like?"; "What do you think you are gaining from them?"; "How have they failed to meet your expectations?"; and "Do the courses flow in a sequence that builds toward student-teaching?" are all examples of interview data that would be valuable. All of these questions could be asked in the last required course before the professional semester, typically in courses called "Methods and Materials" or "Curriculum and Methods." These data would further define the processes of the teacher education program and should be sent to the faculty members for their comments and input.

Another area of interest in process evaluation is attrition data. Students throughout the experience choose to change majors for a variety of reasons. "What are those reasons?" and "What implications can be drawn regarding the teacher education program?" are examples of such questions. Some students change schools; others change because of the financial rewards given in other occupations; others leave because life plans are uncertain. When and why students leave the program are important data for decision-makers as these reasons may reflect upon the program requirements or sequence. Program planners could then identify those alterations. For instance, there may be a problem in

communication with entering students, or the effects of the program sequence on a particular major (e.g., secondary education) may be too demanding or begin too late. The changes made in the program may include improving the advising process or changing the program pattern for some majors. The essential ingredient is that the decision to make changes would be based on systematically acquired student input.

A third area of process evaluation is student opinion of instruction. It would be helpful to know what the students' opinions are of their education courses. The focus should be on objectives, learning activities, field experiences, and course content. These data would not be used to evaluate the instructor but would concentrate on student opinions about the effectiveness of the education courses in meeting selected objectives.

There is also a side effect to student input. Those who write in inservice teacher education recommend that teachers become involved in the planning process of their inservice activities. The use of student interviews indirectly demonstrates to the students that their input is valued and an integral part of evaluation in education. The evaluation process demonstrates to the students that evaluation is not synonymous with inability as it is usually perceived in schools.

The purpose of process evaluation should be to describe the experiences which students and faculty have during the program. The description, although less "hard" than typical product evaluation, does provide decision-makers with information which sets the context in which decisions about the program will be made. The results of process evaluation should lead to informed decision-making, although rarely will decisions be made until the product evaluation data are collected. However, there may be some small adjustments in process and such alterations are not precluded. The next phase is product evaluation.

Product Evaluation

There are numerous products of the teacher education experience and the purpose of this section is to suggest those that are important to evaluation. Typical among the products are cognitive performance, attitude, and behavior assessment. The overriding question in product

evaluation is "What are the students like at the completion of the program?" Specifically, it would be helpful to have tests that measure what the students know about teaching. These would be objective tests that measure the students' knowlege of theory and practice in education. The National Teachers' Exams have attempted, amidst controversy, to meet this need. However, there has been too much debate about the content of the NTE for them to have face validity.

It is foolhardy to want a test that does not match with the objectives of a school's program. The objectives specified by the faculty should dictate the content of the test. However, within that constraint, there are some aspects worth emphasizing as important.

Knowledge competence. As Borich (1979) states,

> "Knowledge competence is the ability to accurately recall, paraphrase, or summarize the procedural mechanics of the behavior on a paper and pencil test."

What the students know about teaching is important. This is measured best by locally-developed and validated objective and essay tests. In order for a test of knowledge to be developed, a model for program evaluation should guide the faculty to:

a) Select or create the objectives of the teacher education program. The model should provide, for instance, the steps in using a delphi technique or other sort systems, by which objectives are derived. Or, the objectives may be drawn from the teaching effectiveness literature.

b) Develop and field-test a measure of what the students know about teaching. An evaluation model should give direction in test writing and item development. A well-conceived test should measure what the student knows about the specified objectives. The scope of the test should include not only recall, but problem solving, critical thinking, decision-making, and the ability to communicate in written form. For example, the test may set up problems and the students must create solutions and strategies for solving them, which draw upon the content of their programs.

Attitude and personality. Product evaluation should also include a post-training measure of any attitude or personality measures that were administered to students upon admission into the education sequence. The changes in the students' scores which have occurred over time would provide another insight into the effects of the program. Few programs know the attitude or personality outcomes of their programs and how student affect may relate to knowledge competence or teaching

performance. A model for evaluation should build in a pre-post design which profiles the changes in student attitude and/or personality.

Teacher performance. A third domain for evaluation, and the one most typically attended to, is performance assessment. One of the remnants of the post-CBTE era is the continued recognition of performance competencies. These are teacher behaviors which have been identified as what preservice teachers should be able to demonstrate in a classroom environment. Medley (1977) has labelled this program evaluation. Program evaluation is the degree to which the training program produces student-teachers who are able to demonstrate the competencies of the program. By observing student-teacher behavior on the competencies, the effects of the teacher education program on teacher performance are quickly gained. Aggregation of the observation data across all students in a given major provides a picture of how the students perform before graduation, thereby demonstrating the effectiveness of the training experiences.

There are many ways that performance can be measured. The purpose in this paper is not to enumerate them, but to suggest what a model should do to lead faculty through the process. An evaluation model should guide the faculty in two areas: competency selection and constructing an observation instrument.

a) Competency Selection. As in the development of knowledge objectives, the evaluation model should provide techniques for deriving the behaviors which serve as the competencies. The scope of the list of competencies will be limited only by the number of areas a faculty wants to consider. The behavior competencies may be drawn from research on teaching or from close examination of the program's requirements.

b) Constructing an Observation Instrument. There are various decisions a faculty must make when specifying teacher performance competencies, that dictate the type of instrument that will be constructed. Some of the decisions about the instrument that need to be made are whether it should be high or low inference, the format of the instrument, and ease of implementation across many supervisors. An evaluation model should build in the appropriate questions the faculty should ask itself about the construction of an observation instrument.

Once the data are collected, the program planners could look for those behaviors on which competence is not demonstrated. These data are

then combined with the information acquired in the process evaluation and the list of objectives generated earlier. The discrepancies between stated competencies and performance are analyzed as they relate to the process evaluation data. The discrepancies can further be analyzed for any relationships with the student characteristics data collected upon entry into the program. The end result of this analysis is that the program planners should be in a position to decide whether the program has met the stated objectives. Where discrepancies between objectives and performance exist, the planners can modify the program to incorporate more instruction to increase students' abilities in those competencies. Further, the relationships between student characteristics, e.g., GPA, SAT, or the attitude and personality measures, may provide some explanation of the variance in the teacher knowledge scores or teacher performance scores.

Reviewing the product evaluation phase, one sees that there are three main foci: knowledge of teaching competence, as measured by paper and pencil tests; performance assessment, measured by direct observation; and the post-training measure of the attitude or personality characteristics. It was also suggested that these three areas would provide a more complete picture of the outcomes of a teacher education program. By collecting these data, program planners can examine the relationship between, for instance, GPA and performance on the test of knowledge, or between performance on the knowledge test and teaching performance, or any of the attitude/personality measures and teaching performance. As the number of students increases, the statistical procedures can include more than correlation, e.g., analysis of various or regression with various dependent variables.

Before moving to the last phase, follow-up studies, it may be helpful to review what has been suggested thus far. At this point in the model, the preservice teacher education program is complete. The program planners have the information to give them an understanding of the students who enter the program, what they experience, and what they have gained as a result. An understanding of the program and a knowledge of the outcomes can then be used to examine the strengths and weaknesses of the program. Informed decisions about the program can then be made.

Follow-Up Studies

The last phase of a comprehensive evaluation plan is the inclusion of follow-up studies. As stated earlier, program follow-up studies should be conducted for three years after graduation. There should be two methods used for collecting these data. The two methods are: a) in situ observation, occurring in the first year of teaching; and b) a mailed survey of teachers' perceptions about their training program and its relation to teaching after two years.

During the first year, follow-up data should be collected by sampling a group of graduates who will be observed using the same observation instrument that was used throughout the program. The reason for collecting observation data during the first year is that one can measure changes over time in teacher performance, and uncover the shifts that occur in practice by comparing student-teaching observations with the follow-up observations. If effective teaching behaviors are not transferred and maintained, then there may be implications for program modification and/or inservice program development. This analysis may be seen as a second measure of program evaluation.

The other area of follow-up studies is a mailed survey to a sample of third-year teachers. The purpose of a mailed survey is to assess the teachers' satisfaction with their training program regarding the tasks they are performing on the job. Providing the items for the survey instrument are the competencies used as the outcomes of the teacher education program. It is important to know the teachers' perceptions on three points: a) the importance of each task or competency in their present positions; b) the degree of difficulty they have in performing the competency; and c) the teachers' ratings of their undergraduate preparation for each task. These three areas would provide a picture of what competencies the teachers find themselves performing, how comfortable they are in performing each, and whether they feel that their program adequately prepared them to use those competencies rated as important.

The data gained from follow-up studies can now be placed along side the evaluation data collected throughout the program. Judgments regarding the effectiveness of the processes of the teacher education

program are made as they relate to the selected competencies and objectives. The follow-up phase described above also provides information about the processes of socialization for the beginning teacher. It gives a view of, not only the immediate outcomes of the teacher education program, but also a view of the maintenance and transfer of the competencies. It records changes in teaching performance of graduates during their first year of teaching and their perceptions about their training program as preparation for teaching. The follow-up studies also identify topics for program development in the continuing education of teachers. The competencies on which there are weaknesses can become inservice education programs.

Some Thoughts About Implementation

One of the more obvious questions about the model suggested throughout this paper is, "How can all this be done?" This is a difficult question which may only be answered in practice. Some thoughts about implementation are offered.

It must be remembered that the role of evaluation is to foster dialogue. In light of this role, evaluation is closely linked to faculty development. The model should place evaluation as a regular function within a teacher education program. This can be done by creating a position and/or office of research and evaluation, or the organization of a standing committee for program evaluation. Evaluation would then become an integral part of the teacher education program and a format for data collection could be designed.

A model that requires such extensive observation of teaching will put some constraints on many faculties. Smaller programs may have some difficulty in conducting so many observations. The model suggested here requires in situ observation as part of student-teaching and follow-up, and it is recommended in this paper that follow-up observation be conducted, at least for a year. If it is necessary, a trade-off in the follow-up study methodology could be made, where well constructed mailed surveys are used instead. However, it seems that a complete picture of the teacher education program, its strengths, and its weaknesses can

only be attained by gathering the information suggested throughout this paper at least once.

In sum, the model called for in this paper is rather extensive. In the best of all possible worlds, it may not be extensive enough, and in the world we live in, it may be too demanding and too consuming of time, money, and human resources. However, between these two positions, a functional model can be built. The model should give teacher educators options for program improvement. The data may not quiet education's critics and maybe it should not. Evaluation should, at the least, provide teacher educators with sufficient data to enable discussion of the strengths and weaknesses of programs and suggest avenues for improvement.

References

Borich, G. D. Three models for conducting follow-up studies of teacher education and training. Austin: Research and Development Center for Teacher Education, The University of Texas, 1979.

Hall, G. E. & Hord, S. M. Water around our ankles, fingers in the hole, where is the life raft? Paper presented at the Fresh Perspectives on the Improvement of Teacher Education Programs, The Ohio State University, College of Education, November, 1981.

McNergney, R. F. & Carrier, C. A. Teacher Development. New York: MacMillan, 1981.

Medley, D. M. Teacher competence and teacher effectiveness. Washignton, DC: AACTE, 1977.

Sandefur, J. T. An illustrated model for the evaluation of teacher education. Washington, DC: AACTE, 1970.

Troyer, M. E. & Pace, C. R. Evaluation in teacher education. Washington, DC: American Council on Education, 1944.

BUILDING PROGRAM OWNERSHIP: A COLLABORATIVE APPROACH TO DEFINING AND EVALUATING THE TEACHER TRAINING PROGRAM

Gary D. Borich

The University of Texas at Austin

BUILDING PROGRAM OWNERSHIP: A COLLABORATIVE APPROACH TO DEFINING AND EVALUATING THE TEACHER TRAINING PROGRAM

Gary D. Borich
The University of Texas at Austin

One of the most sharply focused outcomes of our conference was a convergence of viewpoint concerning the need to engender a meaningful sense of ownership in the teacher training program among all interested parties. An implicit and, from this participant's perspective, major result of our conference was that the problems experienced by institutions in evaluating their training programs are not due so much to a lack of models for evaluation as to the lack of a vehicle for engendering a feeling of program ownership. As was stated by one conference participant, "The problem in my institution is not so much the need for new methods for evaluating the teacher training process but the need for some way to get faculty in divergent fields with divergent interests to care about more than their own limited piece of the training pie."

Those familiar with institutions of higher education will recognize that we often attempt to achieve our training goals in piecemeal fashion, through individual courses taught by faculty from different and often highly autonomous departments. The problem is not one of seeing the "forest" (program) "for the trees" (individual courses) as much as it is one of seeing both the forest and the trees. Seeing one's own individual course contributions in the context of the larger training program may be one of the most significant problems faced by teacher education programs today.

The problem seems even greater when applied to the very special problems posed when the goal is not training generally, but professional training particularly. Here, a trainee does not "commence" with simply a major, but with identifiable skills and practices which allow one to join a professional corp with longstanding rules of conduct. The problem is magnified still further in the case of teacher training in that all of the skills and practices to be taught cannot be derived

within any single setting, field of inquiry, or department. The need for coordination, integration, and synthesis within the training program seems absolutely essential, yet so difficult to attain in practice.

Few would argue that coordination, integration, and synthesis are essential ingredients of a training program. When these ingredients are missing from the teacher training experience, one often notices that training is approached from the perspective of how to achieve the most immediate and concrete objectives first, as opposed to defining what the ultimate goal of the training program should be and how to design the program in such a way as to achieve this goal. This often results in an exclusive focus, or overemphasis, on intermediate or enabling objectives, such as the attainment of discrete skills and concepts taught in individual courses, as opposed to longer term professional practices to which the entire program should be contributing. In these "programs" the training task is reduced to a sequence of attacks on specific problems through the vehicle of individual courses. The problem of coordination, integration, and synthesis between these courses and the program to which they contribute is never addressed explicitly.

What is a Training Program?

Simply defined, a training program is an assembly of related activities designed to bring about outcomes greater than those that could be produced by any of its component parts or set of component parts. In a very real sense a program is greater than the sum of its parts. Its advantage over more disjointed and incremental forms of instruction (e.g., the college major) is that it possesses a level of administrative coordination, instructional integration, and behavioral synthesis that surpasses most other forms of education. It is the relationship among its parts and the relationship of these parts to the whole that allows programs to go beyond simple skills and concepts to build to desirable professional practices indicative of the teaching profession. Knowledge and understandings are turned into skills and abilities and these, in turn, are turned into desirable professional practices for use in the real world. This is a goal of professional

teacher training which is not always shared by other forms of education. It is what makes a program more than simply a sum total of credit hours.

There is considerable diversity in the extent to which individual teacher training programs are, indeed, programs. We can find among these "programs" more or less administrative coordination, more or less instructional integration, and more or less behavioral synthesis. If we agree that programs should be an assembly of related activities brought together for the purpose of teaching knowledge, understandings, skills, and behaviors in ways that build upon one another to produce desirable professional practices, then the problem is one of identifying the means by which coordination, integration, and synthesis activities can be brought about to accomplish this end.

Although many systematic barriers to coordination, integration and synthesis within a teacher training program could be identified, the collective judgment of our conference seems to have repeatedly hit upon the notion of "ownership" as at least one of these barriers. The word "ownership" is not used here of course to denote possession in the entrepreneurial sense but rather as a word chosen to express the related notions of commitment, involvement, and pride. Ownership seems to be an apt word if our goal is to promote coordination, integration, and synthesis--that is, if our goal is to promote the notion of a program.

It would not be difficult to believe that the notion of a program would be unattainable without a concomitant feeling of ownership. Indeed, it would be interesting to observe across a number of teacher training institutions the relationship between degrees of "ownership" and the extent to which these institutions have programs that are more than a sum total of credit hours. If coordination, integration, and synthesis within the training program leads to ownership, then our task is one of finding ways of promoting these intermediate agents which are responsible for program ownership.

Elements of Program Definition

There is a need for an approach to thinking about a teacher training program that (a) brings interested parties together in a

feeling of ownership toward the program, and (b) results in a product that can aid program design, development, and improvement.

One of the most crucial steps in the design, development, and improvement process is the creation of a program definition. Program definition is a careful assessment of the why, what and how of a program. It says why this particular type of program is needed, based on current or foreseen conditions which may be internal to the IHE and/or external to the marketplace. It says what program features will serve to satisfy this need and these conditions. And, it says how the program is to be constructed. A program definition contributes three important ingredients to the meaning of a program:

(1) the reasons why the program is to be created, and which technical, operational, and economic constraints will influence the selection of program activities;

(2) a description of what the program is to be, in terms of the professional activities it must provide and the professional practices it must produce; and

(3) a summary of conditions specifying how the program is to be constructed and implemented. This does not necessarily specify which things will be in the program but, rather, identifies criteria by which these things may later be selected or created.

Even when institutions have made a commitment to program definition, the process of preparing a program definition document is a laborious and frustrating task. This is due, in part, to the intensive interactions among program personnel called for by the process and, in part, to the lack of any organized and systematic method for accomplishing the task. Hence, a coherent, concise and meaningful definition of the program for all to see, to question and to improve is often missing, incomplete, or represents unequally the viewpoints and concerns of those who have a legitimate interest in the program. Lacking a complete definition of the program, program administrators, faculty, and local school personnel will make the missing assumptions and decisions about a program because they must in order to get the program operational or to complete their own individual tasks. Nowhere can this be noted more frequently than when individual faculty design and develop courses they "think" are congruent with the professional practices being sought by program administrators, department heads, and

area coordinators or when program administrators, department heads, and area coordinators develop program goals and objectives they "think" are embodied in the intents of individual faculty. Even when the value of a common program definition is recognized, it takes more than determination for the definition to be realized at the individual course level. This is the first of many stumbling blocks in the process of building ownership.

As every teacher educator knows, it is not uncommon for programs to claim consensus and singularity of purpose at the level of broad program goals and aims but to exhibit diversity, competition and even conflict at the individual course level. It is not unusual for consensus to turn to inter se warfare when ambiguities are scraped away, laying bare the results of years of entrenchment in one's own particular course goals, instructional style and professional agenda. In many cases this situation is exacerbated when global, vague or overly ambitious goals are utilized as integrative organizational mechanisms to achieve consensus. Indeed, goal ambiguity fosters consensus about the purpose and function of the program, and highly ambitious goals are always impressive to "outsiders." But, when ambiguous or highly ambitious goals become the undergirding rationale for the program, it is at the cost of coordination, integration, and synthesis.

One fundamental weakness in current approaches to program definition is an inability to foster a common and detailed understanding of program goals and objectives much less articulate the specific activities and outcomes which logically flow from these goals and objectives. Program definition can best be thought of as "program architecture." Program architecture implements, in good order, the function of a program. Program definition is founded on showing what the functional architecture is, also showing why it is, what it is, and constraining how the architecture is to realize its objectives in more concrete form. In a sense a program definition is the scaffolding on which hangs the conceptual details of the program.

Teacher educators do not tend to think much about the functional architecture of their programs and certainly not in an organized and structured way. I have found that one of the most basic difficulties in defining teacher education programs is the tendency for us not to

realize that an architecture to our programs exists. When it is revealed, the usual reaction is "its just common sense," but as has been remarked for ages, common sense is an uncommon commodity. The need is to structure program concepts with a methodology that communicates the full intent of the program and that allows the program to evolve and to be improved by all parties to the training process.

I believe the analysis of program architecture and the design of program architecture cannot be expressed concisely and unambiguously in natural language. Natural language, in the form of prose, is an effective medium with which to convey our feelings, emotions, intents and goals, but it is not very effective in communicating the complex relationships or interactions of the type so germaine to the notion of a program. A precise, rigorous and easy to use framework is needed to represent the conceptual scaffolding and interrelationships by which programs are arranged.

The form in which the program definition takes shape is important to achieving this precision and conciseness of communication. This form must be able to convey the meaning of a program in minimal space so that a reader can see at once everything which can be said about some aspect of the program. Each program characteristic must be carefully delineated, so a reader can grasp the whole message. A reader must be able to mentally walk through the architectural structure which is portrayed just as blueprints enable a manufacturer to "see" the parts working together. And, finally, there must be a language or syntax to tie the different parts of the program together to form a complete picture of the program's architecture.

This appears to be a tall order, but when done it can yield a program definition that is clear, complete, concise, and consistent. By building and reviewing the program architecture incrementally, while the program is being defined (or redefined), all interested parties can have a voice in directing the process. This "voice," when placed within the context of an approach that explicitly considers the viewpoints and ideas of others begins the process of building "ownership." I will return to how we might specifically build ownership within this framework after a brief description of it.

Constructing a Program Definition

Any approach to program definition should be technically rigorous, but it must also be easy and natural for people to use. Therefore, improved methods of preparing a program definition must include a process for defining a program which includes a definition of people's roles and interpersonal procedures. In addition, there must be a basic understanding of systems and a way of documenting system knowledge in both a rigorous and easy to read format. Notation and documentation must be based upon the idea that the human mind can understand any amount of complexity as long as it is presented in easy to grasp small chunks that are structured together to make the whole. It also must provide for documenting current program activities and decisions in a way that permits for frequent review and revision.

The ideas to be described make several assumptions about the design, analysis, and evaluation of a teacher training program. These assumptions are:

(1) The teacher training program is best studied and improved by first expressing an indepth understanding of what the program is which is sufficiently precise to serve as the basis for program design, development, and evaluation.

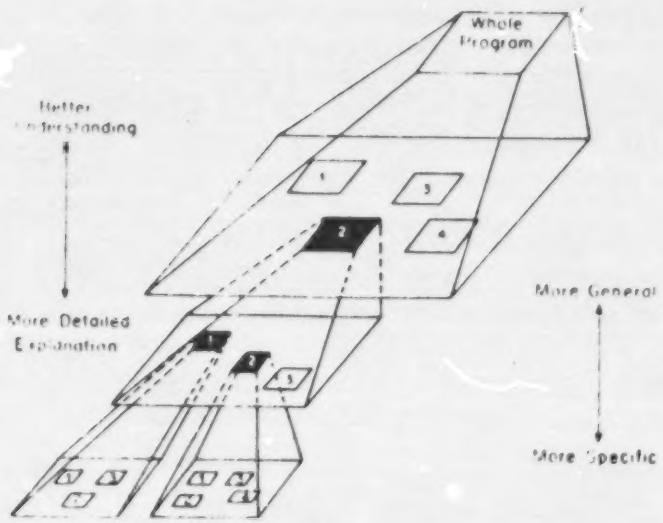
(2) Analysis of the teacher training program should be from the top down, starting with the overall definition of program goals and purposes and moving to specific activities and outcomes which logically flow from these goals and purposes.

(3) Programs are best conceptualized as a series of component parts or individual units with their own goals and objectives but that are structured together to make the whole. At every step of the process the parts being considered must reassemble to make the whole.

(4) A program definition should communicate component parts, their relationships and how they comprise the program architecture. This communication should result in a product that is open for review, revision, and approval by all parties interested in the training process.

Here is how the process might work. Let's start with a general or abstract description of a program to be studied. If we view this description as contained in a "black box," we can break down that box into a number of more detailed boxes, each of which represents a major

function or activity of the program. For ease of understanding let's limit the number of breakdowns (activities) at any step to a small number. Here is what the initial stages of the process might look like on paper. The number of boxes--or program functions--at any given level would of course vary with the nature of the program being studied, as would the number of levels of detail.



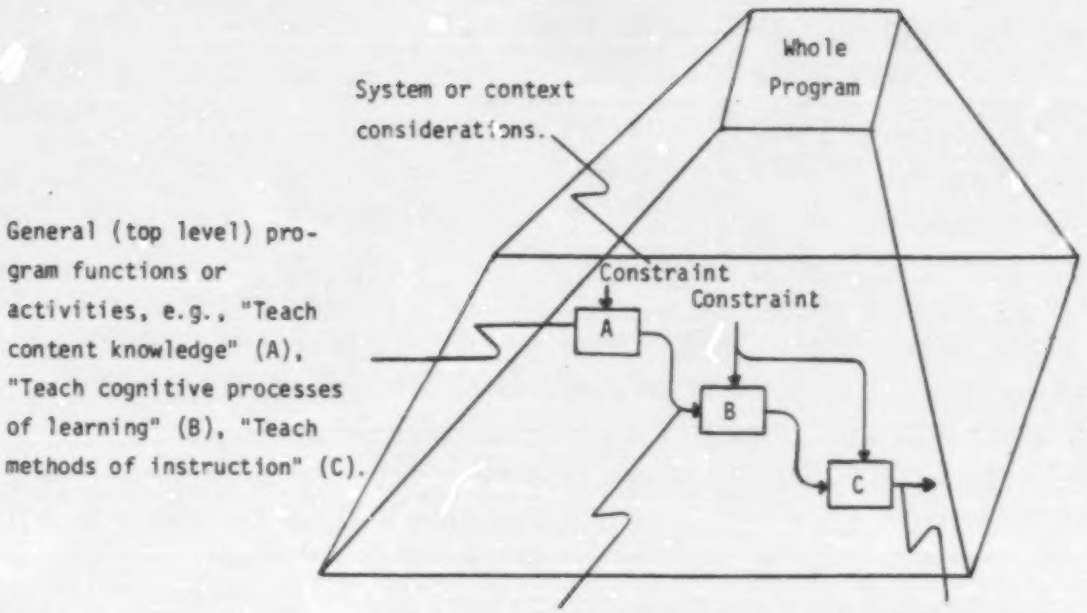
The program is graphically divided into a number of general activities, each symbolizing a major class of events within the program. Each general activity is then further broken down into a small number of subactivities in succeeding steps of the process.

At this point our program definition is an organized sequence of diagrams that describes the functional architecture of the program. A high level overview diagram (top box) represents the whole program. Each lower level diagram shows a limited amount of detail about a well-defined activity. Detailed diagrams clarify and express more specifically the intent and meaning of the "parent" activity. Further, each lower level diagram connects exactly into an upper level diagram to represent the whole program, thus keeping a logical relationship of each program activity to the total program.

Let's further use our graphic technique to show two other aspects of a program: its outcomes and its constraints. As before let's use boxes to indicate program activities or functions. This time we will label them as active verbs in order to communicate exactly what the program is providing. Now, let's connect these activities to the larger outcomes to which they are building by expressing intended relationships between and among program activities. And finally, let's incorporate the notion of constraint to indicate all those context or system considerations that can affect the performance of the program. These context or system considerations are factors within the program or program's environment that "push down" on individual activities to moderate or influence their outcomes. Constraints can be things such as funding, program priorities, feedback on results, prerequisite skills of trainees, motivation of trainers, cooperation of LEA and the like that can influence or moderate the outcomes of any given activity. Since the outcomes of previous program activities can constrain or influence the attainment of subsequent program outcomes, they too can be considered constraints.

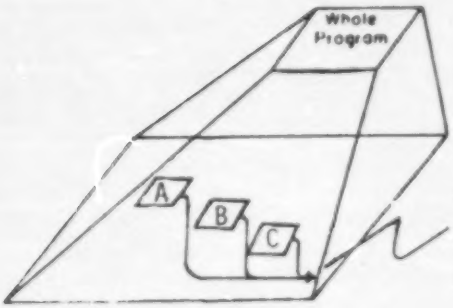
A constraint does not hold back or limit in the usual sense of the word. Rather, it moderates, influences, or dictates what will happen in one of two ways. Constraints can indicate "presses" on a program from somewhere else in the system (e.g., state, IHE, department, LEA, NCATE, etc.) or they can originate at the same level of the system at which the program is operating, (e.g., student readiness for the activity, student skills and aptitude, level of training of program staff, outcome of a previous program activity, etc.).

Here's a sketch of the top two levels of a program when activities, outcomes and constraints are used to communicate a program definition.

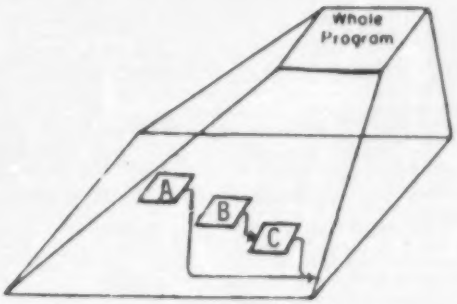

Whole Program
System or context considerations.
General (top level) program functions or activities, e.g., "Teach content knowledge" (A), "Teach cognitive processes of learning" (B), "Teach methods of instruction" (C).
Constraint
Constraint
A
B
C
Specific outcome, tends more toward enabling outcome than longer term professional practice. Outcome of A influences (constrains) B, etc.
Comprehensive outcome, tends more toward professional practice than specific skills or knowledge. It is a consequence of activities A, B, and C.

Although the sequence implied by this representation looks important, its not really. What is important is the dominance implied by showing in "staircase fashion" that the degree to which the outcome is achieved by a trainee at one activity may affect how well he or she will achieve the outcome of another activity. Consider the general belief that the ability to subtract and to multiply must precede the ability to divide. Out of this belief we sequence training such that students can already subtract and multiply before we attempt to teach division. The sequence is thought to be important. It's not really. What is important are the constraints on learning division, two of the most salient of which are the ability to subtract and multiply. Constraint and sequence so often go hand-in-hand that we are cognizant only of sequence, when actually it is the principle of constraint that is most critical to the design, development, and improvement of a program. One implication of this is that sequence does not always imply

constraint. We might ask, for example, to what extent is the sequencing of individual activities in a teacher training program a result of constraint and how much of this sequencing is due simply to convention and tradition. Knowledge of constraint aids our conceptual understanding of a program (and hence its design, development, and evaluation) while knowledge of sequence aids our implementation of a program. Below are two of the many possible alternative designs for a program, communicated by utilizing the notion of constraint.



Terminal outcome, tending toward professional practice.



All activities are independent of one another.

"A" does not constrain "B" or "C" but "B" constrains "C" in order to attain terminal outcome.

This raises an important point. We are not drawing a wiring diagram, flow chart, or organization chart when preparing a program definition. We are attempting to define a program not in terms of the steps or stages required for its implementation but in terms of its instructional activities and the interactions and relationships that make these activities build upon one another to achieve outcomes more comprehensive than can be achieved by any single activity or set of activities. One of the paths to understanding those interactions and relationships is through the notion of constraint.

By linking activities to one another, thinking about constraints both internal and external to the program, and identifying the outcomes which emerge, we avoid ambiguities and vagueries of thought about a program and arrive at the architecture which distinguishes a certain set of activities as a program and not just as a sum total of credit hours.

The approach to program definition I am describing structures a program conceptually. And, it is this conceptual structure or architecture that is most needed for program design, development and evaluation. It is also that which is most ignored in current approaches to analyzing and improving teacher training programs. I have simply introduced the rudiments of an approach for understanding, documenting, revising, and communicating the definition of a program through this architecture.

In short, I have suggested how to prepare a program definition document which allows the meaning of a program to emerge from:

(1) an orderly and well-structured decomposition of intended program activities,

(2) the creation of conceptually defined subunits of program activity that are sized to suit the modes of thinking and understanding of program personnel, and

(3) the representation of program parts which always shows their relation to other parts and to the whole.

This process may be carried out to any degree of depth, breadth, and scope.

The Human Side of Program Definition

When thinking about why program definitions are neither well-structured or well documented, one must not forget that any proposed approach must be people oriented. Behaviors, activities and constraints matter very much, but it is the wishes, ideas, needs, concerns, and skills of people that determine the program. The definition of a training program can only be addressed through the interaction of all people who have an interest in it. One of the current difficulties in preparing a program definition is that individual program personnel often implement their own planning and design concepts. Their background leads them, however well intentioned, to think of program parts rather than program wholes and to measure outcomes and not to define activities or their constraints. A failure to define the program as parts working in concert to produce the whole leads to the very lack of coordination, integration, and synthesis noted earlier as a stumbling block to building ownership. Hence, without a

program definition to bring conceptual clarity and precision to program interactions and relationships there often appears to be no need within the training program to address explicitly the relationship between individual courses and the whole program. This reduces the training experience to a series of one semester attacks on specific behavioral goals. Even when program modifications seem called for by NCATE, state certification boards or professional standards and practices, often a program definition still is not prepared. If modifications are called for in area Y, the program revises or creates a course in area Y. If what is done in area Y creates a problem in area Z, area Z, in turn, is "fixed." Thus, program modifications are confronted through a sequence of remedial attacks on specific problems and the coordination between program areas and the whole program is never explicitly addressed.

Consider a typical set of people who must actively participate in defining the nature of a teacher training program. The customer is the IHE with the need or responsibility for a program. That "customer" authorizes a team to build and maintain a program for a specific population of users. The team probably knows less about the process of program development than it does about the program content that will be implemented. And then there is the state that must certify the program, NCATE which must approve the program, and the LEA's in which program participants must observe, student teach and gain employment. For each of these administrative structures, there is a management group. The program definition must be understood by all of these parties, answer the questions they have about the program, and serve as the basis for shared understanding of the program leading to a feeling of commitment, involvement, and pride.

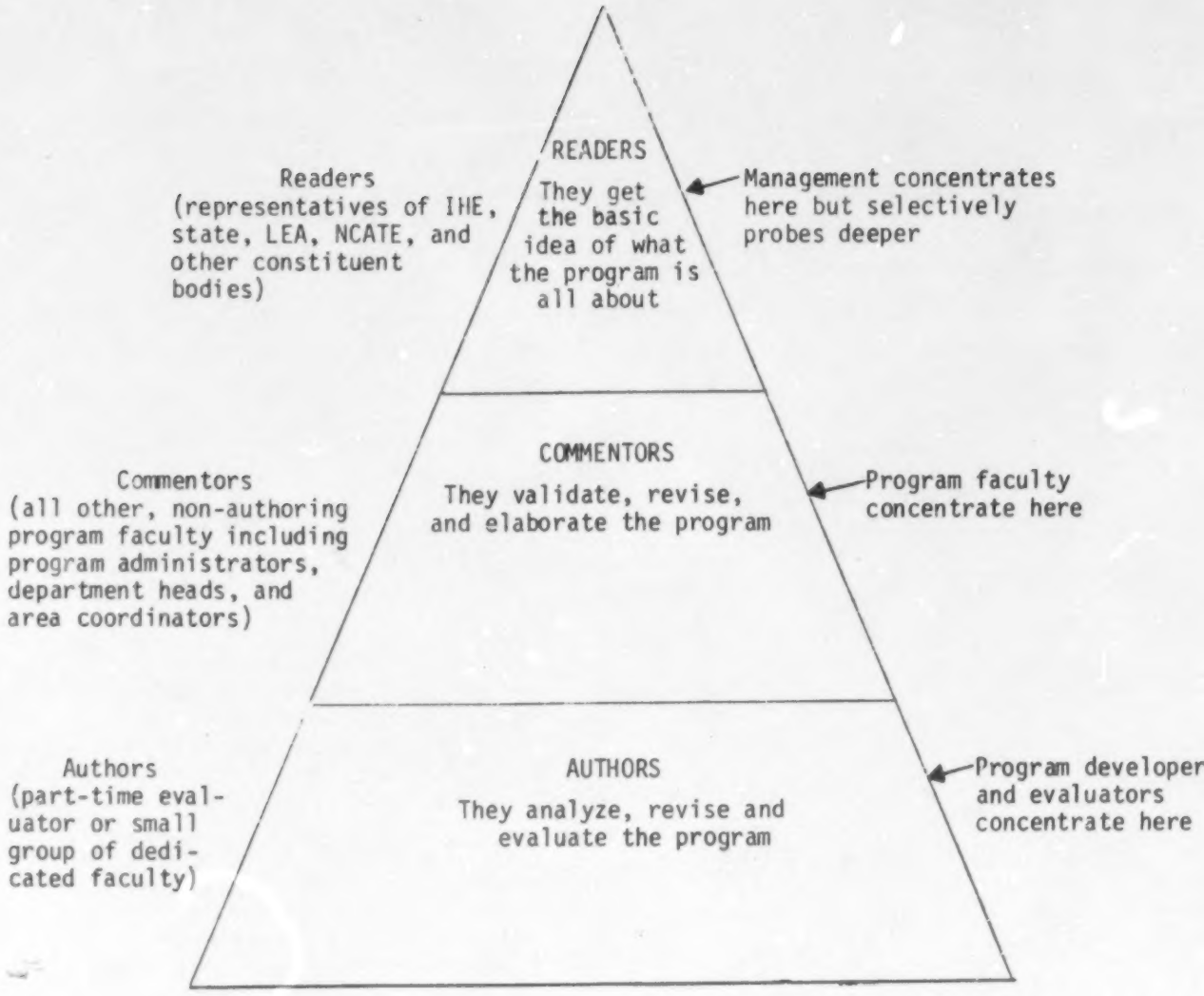
Each of these parties is a partisan whose conflicting, and often vague, desires must be amalgamated through the program. There's a need for individuals to get the assorted information on paper and to structure from it an adequate program definition. The mental facility to comprehend abstraction, the ability to communicate it with personal tact, along with the ability to accept and deliver valid criticism, are all hallmarks of what, hopefully, would characterize these individuals.

These may be part time evaluators, program administrators or a small group of dedicated and concerned faculty or a mix of these. But

as professionals, they are expected to seek out program definitions among other concerned parties. And to succeed, the analysis must be properly managed and coordinated to be receptive to multiple viewpoints. These viewpoints may be overlapping or they may be contradictory. More than likely they will be both overlapping and contradictory.

Throughout the project, draft versions of the graphic program definition should be distributed to other program personnel for review and comment. This approach requires that each person involved in the program will have an opportunity to make comments about the definition in writing. After suggested changes are considered, the updated definition can be reviewed by a committee consisting of representatives from each content area for which the program provides training as well as representatives from the state, the LEA's to which the program is most closely associated as well as the program administration. Such an approval cycle continues upward in the organizational structure until eventually the entire program definition is accepted.

Since the documentation is produced as the definition evolves, the status of the project should become highly visible and this is to be encouraged. Program administrators and area coordinators can study the design in a top-down manner, beginning with an overview and continuing to any level of detail, as can representatives of constituent bodies and the teaching faculty, as shown in the figure below.


READERS
They get the basic idea of what the program is all about
Readers (representatives of IHE, state, LEA, NCATE, and other constituent bodies)
Management concentrates here but selectively probes deeper
COMMENTORS
They validate, revise, and elaborate the program
Commentors (all other, non-authoring program faculty including program administrators, department heads, and area coordinators)
Program faculty concentrate here
AUTHORS
They analyze, revise and evaluate the program
Authors (part-time evaluator or small group of dedicated faculty)
Program developer and evaluators concentrate here

Roles and Functions in the Program Definition Process

Written records are retained of all decisions and alternative approaches as they unfold during the project. Commentors document their suggestions directly on the copies of the diagrams. Authors respond to each comment in writing on the same copy. Suggestions are accepted or

rejected in writing along with the reasoning used. As changes and corrections are made, all versions of the program definition are entered in the project files. A librarian provides filing, distribution and record keeping support, and, not so incidently, also insures precise control over the status of the evolving definition. Since everything is on record, future decisions can reference previous decisions. The following are some of the role players in this process:

Role	Function
Authors (part time evaluator and small group of dedicated faculty representing specific areas of program activity, e.g., reading, social studies, English, science, educational psychology, etc.)	Personnel in respective areas of expertise who study requirements and constraints in their areas, analyze program functions and represent them by producing graphic diagrams of relevant portions of the program, leading to a complete program definition.
Commentors (all other program faculty)	Those who review and comment on the work of the authors. Content personnel who read the diagrams for accuracy and who are expected to make written comments and revisions.
Readers (peripheral program personnel: state, LEA and NCATE representatives, etc.)	Persons from whom authors obtain specialized information about system requirements and constraints usually by means of interviews, telephone calls and review of policy documents.
Constituent Committee (representatives from each constituent body--one each from program areas, and from selected LEA, state, and program administrators)	A group of senior personnel assigned to review the analysis at every major level. They may resolve technical issues or recommend a decision to project management. They may or may not be expected to make written comments.
Project Librarian (secretary, clerical)	A person assigned the responsibility of maintaining a centralized file of all project documents, making copies, distrib-

	uting diagrams for others to see and to comment on.
Project Manager (program director)	Person who has the final responsibility for carrying out the program definition and for preparing the final program definition document.
Instructor	A person comfortable in the approach described, who instructs and acts as a resource for authors and commentors using this approach for the first time.

Summary

I have suggested an approach to (a) thinking in a structured way about a teacher training program, (b) working as a team with division and coordination of effort and roles, and (c) communicating the meaning of a program in clear and concise notation which can aid in program design, development, and evaluation.

How the ideas discussed in this paper are employed can vary according to institutional needs and the kinds of programs under consideration. The process of developing a program definition, however, is as relevant to a new and developing program as it is to an old and established one. This is because its primary goal is to promise interaction and a sense of commitment, involvement, and responsibility on the part of those who apply it. It is not to produce diagrams or a model of the program.

This paper has been based on the premise that a common sense approach to program definition is not now widely appreciated and that a change is needed in the ways that teacher educators think about their programs and about the development and evaluation tasks that they perform. The significance of this approach seems to be that a well-structured approach to documenting what teacher educators think about a program can materially aid both their thinking and their ability to convey their understanding to others. Properly channeled, this understanding can build a feeling of commitment, pride and ownership in

the teacher training process within an institution. But only by considering and understanding a program's architecture or conceptual structure can this commitment, pride, and ownership be brought about. Communication with constituent bodies, the understanding of the nature and structure of programs and a thorough knowledge of the process of program definition may be essential ingredients for the coordination, integration, and synthesis that is necessary for a successful teacher training program.

Finally, some may have noticed that I have tended in this paper to define the work of program involvement as a social process and not as a technology. This reflects my feeling that the process of program design, development, and evaluation has too long been ignored as the human interactive process that it is. In my opinion, the problems of improving teacher training are not so much marked by a need for psychometric instruments, data collection models, or data analysis strategies (which was the theme on which our conference began) but by a need for a better understanding of the conceptual structure of the programs we create. This I contend requires disciplined human interaction for which no amount of measurement technology will suffice.

INTRODUCTORY GUIDELINES FOR DESIGNING EVALUATIONS OF TEACHER EDUCATION PROGRAMS

Nick L. Smith

Northwest Regional Educational Laboratory
Portland, Oregon

INTRODUCTORY GUIDELINES FOR DESIGNING EVALUATIONS OF TEACHER EDUCATION PROGRAMS

Nick L. Smith
Northwest Regional Educational Laboratory
Portland, Oregon

Overview

The purpose of this document is to assist individuals with little or no experience in designing evaluations of teacher education programs to conceptualize the process well enough to carry on meaningful discussions with qualified evaluation practitioners. This report will not enable an educator to do an evaluation, but will assist him or her in planning an evaluation in consultation with an experienced evaluator.

This document contains neither a complete manual of evaluation procedures, nor a discussion of conceptual models for evaluating teacher education programs. Instead, it contains lists of options to be considered in addressing the following five questions basic to the design of such evaluations: What is to be evaluated? Why is the evaluation being conducted? What questions should be asked? What methods should be used? How good is the proposed design?

Each evaluation of a teacher education program must be tailored to fit the local context in which it is to occur, the needs and interests of relevant parties, and the resources available. No universally applicable evaluation design is possible, just as no universally applicable research design exists. Creating an effective design requires the professional judgment of individuals experienced with both evaluation and teacher education programs. This document is designed to facilitate the initial thoughts, discussions, and plans of such individuals.

This document focuses only on the evaluation design process. Other significant elements of the evaluation process, such as the implementing and managing of the evaluation, selecting instruments, analyzing data, and reporting the results are not considered here. Indeed, to treat properly each of these elements would require much more space than is devoted here to the discussion of the five introductory questions.

Five sections are included in this report, each discussing one of the five major design questions stated above. Each section addresses what is to be decided, why the decision is important, what the possible resolutions to the decision are, and how a satisfactory decision can be reached.

Although there appears to be a logical sequence in which these questions might be answered (e.g., evaluate what? why? which questions? what methods? etc.), in practice one cycles through the questions numerous times. An indecision about the appropriate method to use often compels one to reconsider just what the crucial evaluation questions are, which leads one to reassess why the evaluation is being done in the first place. Therefore, there is no assumption made that the document should be read or used from front to back: the sections are designed to be self-contained with references to other sections as necessary.

Section 1: Deciding What to Evaluate

Many elements of a teacher education program might be evaluated. This section helps the reader consider which elements are worthy of attention.

Deciding what to evaluate is a more difficult task than it might first appear because of the large number of possibilities (see below). Such decisions are necessary, however, in order to focus the evaluation efforts and to achieve clarity about the purpose and nature of the evaluation. The decision of what to evaluate influences the subsequent choice of evaluation questions (Section 3) and methods (Section 4) and is influenced by considerations of why the evaluation is being conducted (Section 2). Discussions of what to evaluate often highlight differences in attitudes of program staff, students, researchers, educational administrators, and other audiences, concerning what is important to know about teacher education.

One means of addressing the question of what to evaluate in a teacher education program is to identify the elements that might be scrutinized within the program. Table 1 contains a list of possible program components.

Table 1

Possible Components in a Teacher Education Program

Instruction

- instructional courses - substantive topics
- instructional courses - methods or "professional" topics
- instructional courses - inservice workshops
- curricular materials
- practicum activities
- exams, reviews, proficiency assessment procedures

Goals-Philosophy

- research knowledge supporting instruction
- philosophy or rationale structuring the program
- program goals and objectives

Personnel

- trainees - as students
- trainees - as teachers
- program instructional staff (faculty)
- supervisory teachers/administrators in schools
- program management/administrative staff

Management

- program management/administrative procedures
- program resources - funds, facilities, equipment, etc.

Table 1 identifies 16 components that might appear in a teacher education program. It is possible to identify a greater or smaller number of such components depending on the level of analysis one uses in describing a given program. Further, some components may be missing from certain programs (e.g., programs which consist primarily of a sequence of courses leading to a particular teaching degree) or they may be implicit in others (e.g., programs without formally articulated rationales or goals or without distinct management procedures).

Each of the components listed in Table 1 or any combination of them could serve as the focus of an evaluation. One might begin to decide what to evaluate by selecting a category of components (instructional concerns or personnel concerns) and then proceed to select individual components within that category, or one could select from across categories. An evaluation of a given teacher education program could, therefore, focus on such single topics as:

entry and exit characteristics of trainees as students
adequacy of program goals
quality of curricular materials
effectiveness of practicum experiences
commitment of instructional staff
competence of trainees as teachers
sufficiency of program resources

or on a combination of these components, such as:

the compatibility of the assessment procedures with the program philosophy
the recency of the research knowledge reflected in the curricular materials
the influence of methods course content on trainees' performances as teachers
the resource coordination among program management, instructional staff, and supervising teachers
the administrative procedures for relating practicum activities to substantive course content.

With 16 program components (admittedly an arbitrary number), there are literally thousands of possible combinations (taking the components 1, 2, 3..., etc., at a time) that could serve as the focus of an evaluation.

Additional foci of an evaluation can be generated by considering not just the internal program components, but the program's relationships to external conditions, events, and groups (such as its

relationship to education or to society in general, to its own local context, to the teaching profession, etc.). For example, an evaluation could focus on such topics as:

the program's relationship to other manpower preparation programs within the college
the influence of economic and political changes on the political conditions on the program
the program's contribution to changes in teacher professional identity and militancy.

To fully appreciate the difficulties in deciding what to evaluate in a study of a teacher education program, consider that a large number of questions can be asked about any of the possible foci suggested above (see Section 3).

In light of the large number of possible answers to the question, "What is to be evaluated?" how does one select a limited number of foci for actual study? The selection is best made after a thoughtful consideration of a wide range of information. Table 2 lists some of the sources of information that can be used in considering what to evaluate.

Criteria for selecting from this information and for deciding what to evaluate can be developed from (a) studying the information sources listed in Table 2, (b) analyzing the purpose of the evaluation and the intended use of the resulting information (see Section 2), and (c) considering the nature and quality of the evaluation ultimately desired (e.g., such concerns as feasibility, quality of information, relevance, etc., see Section 5). From this information and criteria it is possible to build a rationale for why a certain study should focus on the particular elements it does.

Although a teacher education program evaluation may include the study of isolated components or of relationships between components, a program-level evaluation will include an assessment of the overall nature and performance of the teacher education program. The importance of global questions should not be overlooked in attending to the quality of individual program components.

Two concerns are especially important in deciding what to evaluate in a teacher evaluation program. The first is to maintain a balance between scope and depth. Generally speaking, choosing a larger number of foci to study results in a more costly, complex, and unmanageable

Table 2

Possible Sources of Information for

Deciding What to Evaluate

- A review of program philosophy, rationale, goals, objectives, and theory-derived expectations
- A review of program materials, activities, and facilities
- A review of past evaluations of the program, and of evaluations of similar programs
- A needs assessment or stakeholder's analysis which identifies what information various audiences want
- A summary of mandates or external demands for evaluative information from the program
- A listing of pressing program issues or identified problems
- An analysis of existing information and research on topics related to program operation
- A review of available resources, time, and staff capabilities
- A summary of staff interests, needs, and curiosities
- An analysis of the state of the art in measurement, statistics, and evaluation of program elements

design, although a greater scope of information and fidelity to actual program reality can be achieved. Choosing a few foci allows for greater depth and a more easily managed study, but at the cost of over-simplifying a complex program and losing the overall picture.

The second concern has to do with maintaining a balance between early design closure and later closure. Selecting evaluative foci early in the evaluation process increases the possibility of providing more precise answers to discrete questions, but at the loss of design flexibility should more important evaluation interests or questions subsequently arise. Delaying longer in selecting evaluative foci enables the evaluation design to be adapted to changing conditions, but at the cost of precision, collection of pre-test or base rate information, and so on. The twin dangers to guard against here are that of providing a precise answer to a no-longer relevant question versus that of providing an inadequate answer to an important question.

Summary of Section 1:

- A clear definition of what is to be evaluated is necessary for an effective study.
- The decision of what to evaluate is tied to the decisions of why to evaluate (Section 2), what questions to ask (Section 3), and what methods to use (Section 4).
- A large number of simple and complex evaluative foci can be identified by examining program elements (Table 1).
- A number of information sources are available (Table 2) in deciding what to evaluate and in constructing a justifying rationale for the decision.
- It is important to balance scope versus depth, and early design closure versus later closure in deciding what to evaluate.

Section 2: Deciding Why to Evaluate

In designing an effective evaluation of a teacher education program, it is important to achieve clarity about why the evaluation is being conducted. This section deals with that issue.

As with the question of what is to be evaluated, the question of why evaluate needs to be addressed, but there will be no simple answers. In designing an evaluation, it is important to be clear about why it is being conducted because its purpose materially affects what is evaluated (Section 1), what questions are deemed relevant (Section 3), and what methods are appropriate (Section 4). In order for a study to fulfill its purpose, that purpose needs to be made clear.

An answer to the why evaluate question will therefore be a statement of the purpose or purposes of the study; that is, the use to which the results will be put. These uses may, of course, be political (to engender public support), managerial (to reduce staff workload) or instructional (to increase the level of skill acquisition).

Being clear about the purpose of an evaluation enables one to more easily answer the other design questions of what to evaluate, which questions to ask, and what methods to use. If an evaluation is conducted to aid an internal decision, then formal evaluation documentation may be unnecessary, but if the evaluation is done to insure program continuation, then full documentation and publicly credible methods are essential.

Table 3 lists some of the possible purposes for conducting evaluations of teacher education programs. Studies are often not conducted with a single, clear purpose in mind, but with multiple purposes which may shift over time. Various audiences will see the evaluation serving differing purposes and will interpret its results in light of those perceived purposes. Thus, while it is common early-on to solicit suggestions from a variety of audiences as to what purposes the evaluation ought to serve, it is important as the study progresses to emphasize repeatedly the purposes it has been <u>designed</u> to serve.

There are at least four ways in which one can get clear about the possible purposes of a study. First, one can discuss the reasons for the study with those who commission or support the study, probing as to their intents, their interests in the work, and the use they see being made of the results. (Be wary of those who commission a study in which someone else is to use the results--no one may actually want or use the information generated.) Second, one can interview program staff, faculty, and school personnel to uncover what purposes they would <u>like</u>

Table 3

Possible Purposes of an Evaluation

- To assess the quality of what has been achieved (summative assessment)
- To solve problems or to refine program operations (problem solving)
- To alter the program into new directions (implementation planning)
- To modify the program in response to changing needs (responsive redirection)
- To maintain managerial or funding agency control (managerial control)
- To obtain accreditation (program accreditation)
- To justify continuation (program continuation)
- To meet an imposed requirement (demand compliance)
- To assure public audiences of program quality (accountability)
- To increase understanding of program processes (research contribution)

the study to serve. (Interviewing those who commissioned the study tells one what purposes it is supposed to serve.) Finally, there are two ways of inferring what purposes the study might serve: by studying past evaluations, current political conditions, and the educational climate, one can infer possible purposes the study might be put to; studying proposed evaluation foci (Section 1) and the related evaluation questions (Section 3) also allows one to infer possible purposes.

Because discussions of study purposes frequently result in people inferring what others' intents are and guessing at others' motivations, it is often more helpful to start with the more concrete issue of what is to be evaluated and then to "digress" to the questions of why the study is being done and what use will be made of its results. (This is the reason this section on evaluation purposes follows, instead of precedes, the section on evaluation foci.)

Summary of Section 2:

- It is important to be clear about why a study is being conducted because the answer affects the other design questions of what to evaluate, what questions to use, and what methods to employ.
- A number of possible purposes for conducting teacher education program evaluations have been identified (Table 3).
- Intended and desired purposes often need to be inferred, but, to be effective, studies must be designed with discrete purposes in mind.

Section 3: Deciding What Questions to Ask

The heart of an evaluation design is the questions asked or the issues addressed. This section considers the range of questions possible in an evaluation of a teacher education program and the methods of generating the questions.

The selection of appropriate evaluation questions to ask or of issues to investigate is necessary in achieving the intended purposes of the evaluation. The selection of methods, instrumentation, and analysis procedures is, of course, also tied directly to the evaluation questions chosen.

There are multiple ways to generate evaluation questions; three of these ways are described below.

Questions Suggested by Program Components

In Section 1 of this report, I suggest identifying program components as one useful way of deciding which elements of a teacher education program to evaluate. Table 1 contains a listing of 16 such components. This same list can be used to create sets of possible evaluation questions. Multiple questions can be generated for any given program component or combination of components. Examples of such questions are contained in Table 4. The more familiar one is with program evaluation, teacher education programs in general, and the specific program to be evaluated, the more easily one can generate a large number of questions merely by studying the program components list.

Questions Suggested by Program Criteria

Criteria for reviewing teacher education programs can be used to generate lists of possible evaluation questions. Sample criteria and evaluation questions are listed in Table 5. Although different criteria may apply to different aspects of the program, the review of general criteria can suggest alternative evaluation questions.

Questions Suggested by a General Search

The most common means of generating evaluation questions is to conduct a general search using the ten sources of information identified in Table 2 of Section 1 (i.e., review of program goals, study of program activities and problems, review of past evaluations, etc.). This approach not only enables one to identify the appropriate range of evaluation questions but also assists in prioritizing the questions generated.

Many more questions can be generated than can be addressed in a given evaluation. It is therefore important to establish a prioritized

Table 4
Sample Questions Suggested by
Program Components

Internal Component Foci - components taken singly or in combinations

Methods Course	• e.g., does the content of the methods course reflect the latest instructional technology?
	• e.g., does the methods course adequately prepare students to handle routine classroom discipline problems?
Methods Course and Substantive Courses	• e.g., are the substantive courses properly sequenced with the methods course?
	• e.g., are the methods taught compatible with the content of the substantive courses?
Methods Course and Practicum Activities	• e.g., do the practicum activities allow full opportunity for trainees to practice the material presented in the methods course?
	• e.g., are trainee practicum experiences used to modify subsequent method course offerings?
Program Goals, Instructional Faculty, and Supervising Teachers	• e.g., are the program goals fully understood by program faculty and supervising teachers and supported by their instructional and supervisory activities?
Management Procedures, Program Resources, Practicum Activities, and Trainees as Teachers	• e.g., are program resources being most effectively managed to insure practicum experiences which foster good subsequent teacher practices?

External Relation Foci - foci which concern the relationships of program components to non-program conditions or events

Management Procedures Teacher Employment	• e.g., how can the program be managed to insure greater coordination with the districts employing the graduating trainees?
Instructional Courses Industrial Inservice	• e.g., what current program pre-service instruction can be adapted from industrial inservice use?
Instructional Courses Other Instructional Programs	• e.g., how can the program's training capability be strengthened by collaborating with other college instructional programs?

Table 5

Sample Questions Suggested by Program Criteria

Criteria	Questions
Need -	e.g., is there a need for the kinds of trainees produced by the program?
	e.g., is the placement of trainees in practicum positions adequately meeting their field exeperience needs?
Quality -	e.g., what is the quality of the instructional materials used to instruct trainees?
	e.g., what is the quality of the instructional technology equipment used in the program?
Utility -	e.g., how useful are the methods courses in preparing trainees for actual teaching?
	e.g., how useful is the direction provided by the supervising teachers?
Efficiency -	e.g., is the program well managed and overall cost efficient?
	e.g., are the most cost effective means used to provide instruction in high technology?
Support -	e.g., does the program have the support of other instructional units in the college?
	e.g., are districts which hire program trainees generally supportive of the program's operation?
Equity -	e.g., does the program stress equity through the use of non-biased curricular materials?
	e.g., does the program promote equal educational opportunity through balanced trainee selection and placement procedures?
Impact -	e.g, is there evidence that graduates of the program are significantly increasing the quality of education in the districts where they are employed?
	e.g., is the instruction provided by the program faculty significantly increasing the teaching skills of trainees?

Criteria		Questions
Responsiveness	-	e.g., are the program course offerings sufficiently responsive to the changing needs and demands of public education?
		e.g., are faculty members responsive to the special needs of individual trainees?
Potential	-	e.g., does the program prepare trainees for leadership positions in the educational enterprise?
		e.g., does the program show potential for continued growth and increased community service?

list which can be used to guide the selection of evaluation methods and the allocation of evaluation resources.

Three sources of information in Table 2 are especially important in prioritizing evaluation questions: pressing program issues or problems, mandates and external demands for information, and the identification of audience information needs through needs assessments or stakeholders analyses. In identifying the information needs of audiences or stakeholder groups, it is important not to overlook the variety of possible audiences. A summary of possible audiences for teacher program evaluations are listed in Table 6.

The final prioritization of evaluation questions is, of course, heavily influenced by the purpose of the evaluation (how the results will be used, Section 2) and by the focus of the evaluation (the "what" to be evaluated, Section 1).

Summary of Section 3:

- The specification of evaluation questions is one of the most crucial aspects of creating a good evaluation design.
- There are many ways to generate evaluation questions, including examining program components (Table 4), reviewing program criteria (Table 5), and conducting a general search (Table 2).
- Prioritizing possible evaluation questions is essential, and analyzing audience information needs is a useful strategy for determining priorities. Care must be taken, however, to consider all relevant evaluation audiences (Table 6).

Section 4: Deciding What Methods To Use

It is best to make decisions about which evaluation methods to use after one is clear about: the purpose of the evaluation and the use to which the results will be put (Section 2), the foci or program elements to be evaluated (Section 1), and the evaluation questions or issues to be addressed (Section 3).

Methods which are compatible with criteria such as those listed in Table 7 should then be selected.

The term "program evaluation" generally refers to the assessment of the overall programmatic effects of a social or educational program.

Table 6

Possible Audiences for Evaluations of

Teacher Education Programs

- Evaluation Sponsors/Funders
- Program Sponsors
- Program Faculty
- Program Administrators
- Program Trainees
- Program Graduates
- Supervisory Teachers
- Employing Administrators
- Educational Researchers
- Other College Faculty and Administrators
- Other Teacher Training Program Faculty and Administrators
- Teachers
- Principals and School Superintendents
- Teacher Union Officials
- School Board Members
- Non-College Inservice Trainers
- State Boards of Education
- State Licensing Personnel
- State and National Accreditation Personnel
- Students
- Parents
- Special Interest Groups
- Public

Table 7

Suggested Criteria for the Selection

of Evaluation Methods

Evaluation methods should be selected on the basis of their:

- Fidelity to the purpose of the evaluation
- Suitability to the program components being evaluated
- Relevance to the evaluation questions being addressed
- Utility in providing the needed information
- Quality of information produced
- Acceptability to evaluation methodologists
- Credibility to evaluation audiences
- Compatibility with the setting in which they are to be used
- Flexibility in adapting to unanticipated design changes
- Cost Efficiency in time, resources, and staff capability
- Marginal Utility in relation to other methods
- Non-Reactivity with program components
- Lack of Negative Side-Effects

However, as illustrated in Sections 1 and 3, it is impossible in a single study to evaluate all global and detailed questions that could be generated about every possible combination of program components--hence selecting and prioritizing of questions are essential. Further, the selection and prioritization process may have significant methodological implications. For example, if the primary questions to be addressed concern the quality of instructional materials, then curriculum evaluation methods become most relevant in conducting the evaluation. Table 8 contains a listing of the most relevant evaluation methods for selected sets of program foci.

A great variety of standard and innovative research and evaluation methods are available for use in evaluating teacher education programs. A full discussion of such methods is beyond the scope of this paper, but two illustrations follow. The first illustration suggests the range of methods applicable to a given problem, the follow-up assessment of program graduates (and includes most of the major field study methods currently in use). The second illustration suggests the possible application of a variety of new methods to the evaluation of teacher education programs.

Although follow-up assessments of teacher training program graduates are usually conducted by sending mail questionnaires, a wider range of other methods is available. Examples of these methods are identified in Table 9.

New evaluation methods which can be used for evaluating teacher education programs are becoming available. A sample of such methods is identified in Table 10. While some of these methods may be too esoteric for general use, evaluators of teacher education programs should be aware of the existence and possible utility of these methods. References for the methods identified in Table 10 are included in Appendix A.

Summary of Section 4:

- The use of a set of criteria for selecting evaluation methods is recommended (Table 7).
- Depending on the foci of the evaluation being conducted, methods from

Table 8

Sources of Most Relevant Evaluation Methods for Selected Program Foci

If the evaluation concerns the quality of:	Then the most appropriate methods are found in:
• instructional materials	• curriculum evaluation
• program trainees	• student assessment
• graduate performance	• performance assessment
• program faculty	• faculty evaluation
• program management staff	• personnel evaluation
• program effects	• impact evaluation
• program financial management	• financial auditing
• program philosophy, rationale, and goals	• policy evaluation
• program management procedures	• management evaluation
• program activities	• process evaluation
• program equipment	• product evaluation
• the interaction of these elements	• program evaluation

Table 9

Selected Methods of Conducting Graduate Followup Studies

- Mail questionnaire studies of current activities and evaluative judgments of quality of professional training
- Collection of peer and supervisor ratings of the adequacy of preparation, current performance, and inservice needs.
- Study of classroom management skills through pupil judgments on classroom climate inventories
- Assessment of teaching competencies through classroom observation techniques
- Development of case studies of teacher acculturation problems through use of teacher self-report logs and phone interviews
- Review of teacher instructional development skills through expert critique of teacher-made curricular materials, tests, and plans
- Study of teacher competencies through post-graduation performance testing
- Assessment of inservice training needs through review of school records on teacher performance and promotion
- Demographic studies of graduate mobility and tenure in profession
- Review of interpersonal skills through phone interviews with parents of pupils
- Comparative salary studies using data from state and local teacher unions to assess relative financial success of graduates
- Assessment of continued professional development through unobtrusive studies of graduates' inservice attendance, graduates' use of instructional media facilities and college job referal services, and graduate enrollment in continuing education courses, etc.

Table 10

Sample of Alternative Teacher Education

Program Evaluation Methods

Method	Possible Application
Cost-Effectiveness Analysis (Economics)	to study the cost utility of having two semesters of methods courses instead of one semester
Social Area Analysis (Epidemiology)	to identify the characteristics of teachers needed within a given geographic region
Trend Surface Analysis (Geology)	to study the regional distribution of program graduates
Integer Programming (Operations Research)	to determine how to cut program budgets in a way that will maintain the highest possible quality of the educational program
Transportation Models (Operations Research)	to determine the most desirable assignment of trainees to practicum experiences
Document Tracking (Journalism)	to reconstruct the flow of educational activities and events of past program graduates
Concept Analysis (Philosophy)	to clarify the meaning of key terms, such as "proficiency," "competency," "remedial," "discipline," etc.; a clear understanding of them is crucial to effective teaching
Photography (Photo-Journalism)	to document the nature of trainee-pupil interaction during practicum experiences
Committee Hearings (Congressional Investigations)	to obtain structured public discussion of the merits and deficiencies of program operations
Thematic Analysis (Poetic Criticism)	to uncover the underlying themes or hidden curricula being taught to trainees through the program

areas other than program evaluation may be more appropriate (Table 8).

- There are many possible ways to conduct graduate follow-up assessments (Table 9), and a variety of new evaluation methods is available for use in teacher education program evaluations (Table 10).

Section 5: Deciding How Good the Design Is

Decisions about what to evaluate (Section 1), why to evaluate (Section 2), what questions to ask (Section 3), and what methods to use (Section 4) provide the basic structure for the evaluation of a teacher education program. Much remains to be decided, of course, before a fully detailed design can be prepared. Table 11 contains a list of the possible elements of a complete evaluation design.

It is beyond the scope of this document to provide a complete guide for designing evaluations of teacher education programs--that would require a book, not a brief document. The intent of this document is only to assist in the initial conceptualization of such a design so that more productive planning sessions are possible between program personnel and qualified evaluators.

As a final design is prepared, it is useful to keep in mind a set of criteria for reviewing the design's overall adequacy. The most complete set of criteria developed to date appear in Standards for Evaluations of Educational Programs, Projects, and Materials (by the Joint Committee on Standards for Educational Evaluation, New York: McGraw Hill, 1981). For the reader's information, Table 12 contains a listing of the titles of the thirty standards. Of course, the complete volume, with its full discussion of each standard, including sample applications, should be consulted as the evaluation design is produced.

Following are a few additional points to keep in mind as one plans an evaluation of a teacher education program.

- Utilization research suggests that, in order for evaluative information to effect program change, such information must be locally relevant, be responsive to a felt need, be timely, and be supported by internal program advocates. Faculty involvement and support of the evaluation is crucial. Externally imposed evaluations are less likely to effect significant program changes.

Table 11

Possible Elements of An Evaluation Design

- The purpose of the evaluation (use of results)
- The program foci to be evaluated and the rationale for selecting them
- The questions or issues to be addressed
- The methods to be used

- The major audiences of the study
- The information sources, including sampling plans
- The instrumentation to be used
- The timing and amount of data to be collected
- The analysis steps
- The reporting procedures

- A timeline for the study
- A list of management activities and checkpoints
- A statement of protections and assurances (protection of privacy, informed consent, etc.)
- A staffing plan
- A budget summary

Table 12

The Standards for Evaluations of Educational Programs, Projects, and Materials*

A. Utility Standards

1. Audience Identification
2. Evaluator Credibility
3. Information Scope and Selection
4. Valuational Interpretation
5. Report Clarity
6. Report Dissemination
7. Report Timeliness
8. Evaluation Impact

B. Feasibility Standards

1. Practical Procedures
2. Political Viability
3. Cost Effectiveness

C. Propriety Standards

1. Formal Obligation
2. Conflict of Interest
3. Full and Frank Disclosure
4. Public's Right to Know
5. Rights of Human Subjects
6. Human Interactions
7. Balanced Reporting
8. Fiscal Responsibility

D. Accuracy Standards

1. Object Identification
2. Context Analysis
3. Described Purposes and Procedures
4. Defensible Information Sources
5. Valid Measurement
6. Reliable Measurement
7. Systematic Data Control
8. Analysis of Quantitative Information
9. Analysis of Qualitative Information
10. Justified Conclusions
11. Objective Reporting

---

*From Joint Committee on Standards for Educational Evaluation, Standards for Evaluations of Educational Programs, Projects, and Materials. New York: McGraw Hill, 1981.

- The most effective evaluations are those that can become institutionalized within the teacher education program, that make the maximum use of existing data to effect program change, and that foster coordination across relevant agencies (e.g., local school districts, state departments of education, etc.).

- An important function of an ongoing evaluation effort is to maintain communication across similar teacher education programs in order to identify practices that work, new approaches, experiences with new measures, and so on.

Summary of Section 5

- A list of possible elements of a complete evaluation design is provided (Table 11).

- It is recommended that one use a set of standards for reviewing the evaluation design as it is produced (Table 12).

I wish to thank the following reviewers who provided valuable comments on an earlier draft: Gary Borich, Betty Dillon-Peterson, Jim Leary, James Raths, and J. T. Sandefur. Remaining limitations in the document are, of course, solely my responsibility.

APPENDIX A

References for Alternative Teacher Education Program Evaluation Methods (Table 10)

Cost Effectiveness Analysis

Levin, H. M. Cost analysis. In N. L. Smith (Ed.), New techniques for evaluation. Beverly Hills, CA: SAGE Publications, 1981, pp. 13-70.

Levin. H. M. Cost-effectiveness analysis in evaluation research. In E. Struening and M. Guttentag (Eds.), Handbook of evaluation research. Vol. 2., Beverly Hills, CA: SAGE, 1975.

Social Area Analysis

Smith, N. L. Techniques for the analysis of geographic data in evaluation. Evaluation and Program Planning, 2, 1979, 119-126.

Trend Surface Analysis

Smith, N. L. Techniques for the analysis of geographic data in evaluation. Evaluation and Program Planning, 2, 1979, 119-126.

Integer Programming

Wholeben, B. E. Multiple alternatives for educational evaluation and decision-making. Paper and Report Series No. 72. Portland, OR: Research on Evaluation Program, Northwest Regional Educational Laboratory, May, 1982.

Transportation Models

Caulley, D. N. The use of assignment and transportation models in evaluation. Paper and Report Series No. 68. Portland, OR: Research on Evaluation Program, Northwest Regional Educational Laboratory, February, 1982.

Document Tracking

Guba, E. G. Investigative journalism. In N. L. Smith (Ed.), New techniques for evaluation. Beverly Hills, CA: SAGE Publications, 1981, pp. 167-262.

Caulley, D. N. Document Analysis in Program Evaluation. Paper and Report Series No. 60. Portland, OR: Research on Evaluation Program, Northwest Regional Educational Laboratory, July, 1981.

Concept Analysis

Gowin, D. B. Philosophy. In N. L. Smith (Ed.), Metaphors for evaluation: Sources of new methods. Beverly Hills, CA: SAGE Publications, 1981, pp. 181-209.

Caulley, D. N. Concept analysis in evaluation. Paper and Report Series No. 61. Portland, OR: Research on Evaluation Program, Northwest Regional Educational Laboratory, July, 1981.

Lane, C. A. Using the Tools of Philosophy in School Program Evaluation: Metaphor in Action. Paper and Report Series No. 55. Portland, OR: Research on Evaluation Program, Northwest Regional Educational Laboratory, February, 1981.

Photography

Templin, P. A. Still photography in evaluation. In N. L. Smith (Ed.), Communication strategies in evaluation. Beverly Hills, CA: SAGE, 1982.

Committee Hearings

Stenzel, N. K. Committee hearings as an evaluation format. In N. L. Smith (Ed.), Field assessments of innovative evaluation methods. San Francisco, CA: Jossey-Bass, 1982.

Stenzel, N. K. Committee hearings. In N. L. Smith (Ed.), Communication strategies in evaluation. Beverly Hills, CA: SAGE, 1982.

Thematic Analysis

Della-Piana, G. M. Literary and film criticism. In N. L. Smith (Ed.), Metaphors for evaluation: Sources of new methods. Beverly Hills, CA: SAGE Publications, 1981, pp. 211-246.

Della-Piana, G. M. Film criticism and micro-computer courseware evaluation. In N. L. Smith (Ed.), Field assessments of innovative evaluation methods. San Francisco, CA: Jossey-Bass, 1982.

EXPLORATIONS IN THE EVALUATION OF TEACHER EDUCATION

Daniel L. Stufflebeam

Western Michigan University

EXPLORATIONS IN THE EVALUATION OF TEACHER EDUCATION

Daniel L. Stufflebeam
Western Michigan University

The April 1982 symposium in Austin surfaced and addressed several key questions concerning the evaluation of teacher education. From my perspective, the most important of these questions were as follows:

1. What fundamental conceptualization should guide attempts to evaluate teacher education?
2. What models are appropriate and useful for applying the recommended conceptualization?
3. How should given models be operationalized?
4. What can be done to improve the utilization of evaluations?
5. How can departments of teacher education institutionalize evaluations of their programs?
6. What criteria should be used to guide and assess evaluation work (including that conducted by NCATE), and how should these criteria be applied?

In this paper I will address the first three of these questions by interpreting them, reacting to what I learned from the symposium, and offering a few suggestions. I will offer my advice within the context of my previous writings on the CIPP Model for evaluation. I will attempt to extend my views by referring to the papers presented at the symposium and to illustrate them by referring to problems in evaluating teacher education. Basically, this paper is a stand alone presentation, but the readers' understanding of the paper would be enhanced by their review of the working papers presented at the symposium. These papers were presented by Erly, Galluzzo, Gardner, Roth and Sandefur. Additional background papers were made available by Borich and Craig (see references).

I am indebted to the organizers of the symposium, to the main presenters, and to the other participants who engaged in in-depth discussion of the presentations; their inputs assisted me greatly in the preparation of this paper. Nevertheless, I assume full responsibility

for the particular positions advanced in what follows. I urge the readers to view my remarks as a working paper. I hope it may stimulate thinking and communication about the important area of "teacher education evaluation." I will welcome critical reactions and suggestions so that I may abandon unproductive ideas and improve my views and advice about how best to evaluate teacher education.

1. How should evaluation of teacher education be conceptualized?

If evaluation efforts are to be of service to teacher educators, these efforts need to be guided by a defensible conceptualization. Otherwise, the evaluations will serve other than useful purposes, address the wrong questions, provide feedback too late to do any good, etc. In essence, the guiding conceptualization should clarify the meaning that is ascribed to the term evaluation. I found several notions advanced at the symposium very useful for developing a defensible definition of evaluation. I will review these ideas, then use them to refine a definition of evaluation that I have previously proposed. One important point emphasized in group discussion at the symposium was that evaluation must be viewed as a process, not a product. This is particularly applicable to higher education, since much of the evaluation tradition in this sector has been limited to responding to periodic requirements for evaluation emanating from accrediting bodies. Often this response mode has resulted in a stilted, ritualistic approach to evaluation. If evaluation is to fulfill its potential contribution to assuring the quality of teacher education, it needs to be conceived and implemented as an ongoing program. Only then will faculty, administrators, and monitors be provided with ongoing formalized feedback by which to modify programs.

Another theme running through the symposium, that I readily endorse, is that the main purpose of evaluation is not to prove but to improve. As Roth argued, "The most important reason (to evaluate) should be in order to improve programs and the quality of their graduates." This position counteracts the view that evaluations should be witch hunts or only instruments of accountability. Instead, it sees evaluation as a tool by which to help make programs work better for schools and teacher trainees they are intended to serve. This position

is consistent with those recently presented by Patton (1978) and Cronbach and Associates (1980). I hasten to add, however, that an improvement orientation need not and should not avoid the likelihood that some teacher education programs are unworthy of efforts to improve them and thus should be terminated. By promoting the demise, or at least the decertification, of unneeded or hopelessly flawed programs, evaluations also serve an improvement function. They do so by helping to concentrate the nation's scarce resources for teacher education in a reasonable number of high quality programs.

Two other purposes for evaluation, in addition to improvement, were also reflected in the discussions and papers of the symposium. The need to provide information for accountability was evident in the papers by Gardner, Roth, and Erly. And, Borich argued effectively that evaluations of teacher education should both draw from and contribute to research on teaching. While I recommend that most emphasis be given to the improvement function, I agree that evaluations should be planned and carried out to support accountability needs and to promote understanding of teaching and teacher education.

Given the states' responsibility for education, given the accrediting mechanisms sponsored by professional organizations, and given the need of federal agencies to assure that their financial assistance to teacher education is used appropriately and effectively, teacher education programs inevitably will be required to account for their effective use of resources. To meet these requirements teacher educators need to identify their main audiences for accountability information (e.g., the central administration of their university, their state department of education, and NCATE). Then they should project, as accurately as they can, the information desired as well as required by these groups. They should make sure that pertinent information gathered in their program improvement efforts is filed for later use in accountability reports, and, to the extent feasible, they should gather any additional information that may be needed. Beyond satisfying external requirements, evaluations that are geared to serve accountability (as well as improvement) have the added advantage of helping to develop credibility for teacher education programs.

With some foresight and planning, evaluations can also contribute to improved understanding of teacher education. Given the short supply of funds for research on teaching, exploitation of this opportunity is certainly in the best interest of advancing the science of teaching. Nevertheless, this third purpose must be subserviant to the improvement and accountability functions; for experience has shown that questions of interest to researchers often are much too narrow in scope to meet the evaluative information requirements of program staff and sponsors.

This point leads to another pertinent topic of discussion at the symposium--that concerned with what main questions should be addressed in the evaluation of teacher education. The dominant position, especially as reflected in the papers by Gardner, Roth, Sandefur, and Borich, was that primary attention should be given to assessing the competence and on-the-job performance of graduates of a program. Certainly, an evaluation of a teacher education program that ignored the quality of outputs would be incomplete, and, as pointed out repeatedly at the symposium, past evaluations of teacher education have often failed in this regard. But, as emphasized in the paper by Galluzzo, sound and useful evaluations of teacher education must address a much broader range of questions than those associated with student competence and on-the-job performance. In general, I agree with his view that the evaluations should address questions that pertain at least to admission, process, product and follow up. The full range of such questions must be considered when the primary purpose of evaluation is to foster improvement. Whereas accountability audiences may be most concerned about the competence and performance of graduates, program administrators and staff must also closely examine those aspects of their programs that help to determine success or failure.

Much of the discussion at the conference endorsed the use of clearly stated teacher competencies and program goals as bases for identifying and assessing outcomes. This position is consistent with the longstanding Tylerian tradition that has equated evaluation to the process of determining whether valued objectives have been achieved. I have no quarrel with this recommendation as one approach to obtaining and interpreting outcome data. But it is far too narrow an approach to account for all the data gathering and interpreting that needs to be

done in evaluating teacher education. As Sandefur suggests, the goals themselves must be evaluated, and as Craig prescribes, the competencies must also be validated. But even more fundamentally, evaluations of teacher education need to be oriented to the needs of the teacher trainees and of the schools and society where they will serve. Needs assessment, then, is a crucial part of a sound program of evaluation. Also, the people who are closely associated with any program potentially are the best source of intelligence about the program; hence as Stake (1977) has long argued, we should systematically seek out, analyze, and report their judgments and recommendations.

Another problematical aspect of conceptualizing evaluation concerns what steps are required to ensure both the collection and utilization of sound information. Sandefur called attention to this problem area when he lamented the lack of impact of the extensive evaluation of teacher education performed over a nine year period at Western Kentucky University. His comments provide an excellent overview of what is at issue here.

> "Although the model successfully evaluated graduates, albeit a complex, unwieldy, and expensive process, we were never truly successful in feeding the evaluative data back into the development of our ongoing teacher education program, at least not to the extent we had hoped. Although it must have been our fault, our faculty seemingly never understood the relationship between the data we collected and the ongoing revision of the teacher education program. Perhaps the fault lay in their lack of feeling of involvement and ownership. They seemed to think of the project as the Dean's or Ronald Adams' source of research and publications. Had we the opportunity to conduct the project over, we would have made provision for faculty involvement both in the decision-making and conduct of the project."

In this statement, evaluation is revealed to be both a communication and technical process. I cannot overemphasize my strong conviction that the methodology of evaluation must provide equally for the effective and ongoing communication and collaboration of audience and evaluator and the collection and processing of sound data. Time and again, evaluations that have excelled in either to the exclusion of the other have failed.

Even the preceding cryptic review of the symposium's coverage of conceptual problems in evaluation reveals that arriving at a sound guiding rationale is both a crucial and complex task. Surely, no single statement can adequately account for all the complexities. But I believe that a carefully constructed definition can help to move us away from narrow views of the past (e.g., those that equate evaluation to testing, to comparing outcomes to objectives, or to conducting field experiments) and can point us in the right direction.

In closing my analysis of the first question I have, therefore, decided to propose a definition of evaluation. It is one that I have been working on for a number of years, and I have adapted it to take account of the points I gleaned from the symposium. The proposed definition is as follows:

> Evaluation is the process of delineating, obtaining, and applying descriptive and judgmental information concerning the worth and merit of some program's goals, design, implementation, and impacts in order to promote improvement, serve needs for accountability, and foster understanding.

This definition is complex, but the preceding discussion should provide some insight into its main elements. The definition sees evaluation as an ongoing process. The process includes the technical activity of obtaining the needed information and the communication steps involved in determining what information is needed and applying this information once it is obtained. The word applying was chosen purposefully over other possibilities, such as providing or reporting, in order to emphasize Sandefur's point that evaluations should be collaborative efforts of evaluators and clients to obtain and use data to guide development. The definition provides both for describing programs and program contexts and for collecting judgments from persons with some knowledge and/or concern for the program. The definition reflects Galluzzo's concern for addressing a wide range of questions about a program (in my words those pertaining to goals, design, implementation, and impacts).

The definition also reflects the important arguments advanced by Scriven (1980) and by Guba and Lincoln (1979) that evaluation should examine questions of worth and of merit. While the symposium papers and

discussions did not evidence concern about this distinction, I see it as especially relevant to teacher education. Worth refers to the need within the broader society for the functions served by an institution and its programs. For example, a teacher training program may rate relatively high or low on worth depending on the society's need for the quantity and types of teachers that the program prepares. Again, we see the importance of needs assessment in program evaluation. Merit, on the other hand, is context free and concerns the degree of excellence with which the program prepares teachers. For example, a program would rate high on merit if it excelled in preparing competent social studies teachers, even though the society might not have a need for such teachers. It is important, then, to examine both worth and merit--and not to confuse the two--in order to assess the extent that a program excells (the merit criterion) in meeting important needs (the worth criterion).

Finally, the proposed definition lists these purposes for an evaluation--improvement, accountability, and understanding. To reiterate, I have listed three purposes in what I see as their relative order of importance.

2. <u>What models are appropriate and useful for applying the recommended definition?</u>

The word model was used somewhat loosely at the symposium. I am sure that the different users of this term had different meanings in mind, and, no doubt Kaplan and other philosophers of science would have cringed at our loose use of the term. When, between formal sessions, I mentioned this problem to Gary Borich, he gave me a paper he had written on models and persuasions. In this paper he argued that models in evaluation do not (and perhaps need not) prescribe explicit data requirements and evaluation procedures but, instead, "...are only heuristics which help organize one's thinking about an evaluation..." This point helped to defuse my concern and is the sense in which I will use the term model in this paper. It will allow us to look at the broad array of "models" referenced in the symposium papers and discussions. It will also focus attention on the important question concerning what

models are appropriate and useful, i.e., heuristic for applying the definition of evaluation that I have proposed.

By my count, fourteen models were introduced at the symposium. Gary Borich presented three models by which to conduct follow up studies. Sam Craig discussed six program evaluation models under investigation in Pennsylvania. Bob Roth described what he sees as the needed data base for evaluations of teacher education, particularly, but not solely, at the state level. Bill Gardner described NCATE's process of evaluation for accrediting purposes. J. T. Sandefur reviewed and critiqued the evaluation model used by him and his colleagues at Western Kentucky University. And Gary Galluzzo described the evaluation model employed at Glassboro State College which he said is based partially on the CIPP Model for Evaluation (which I developed). In addition, Nancy Zimpher provided me with a description of the Ohio State University System for Documenting and Evaluating the Experiences of pre/inservice teachers.

An indepth discussion and critique of these models is beyond the scope of this paper, but I would see this as a worthwhile future step. The models do provide an interesting array of alternatives. Some of the differences are due to differences in objects to be evaluated, e.g., total program versus the performance of graduates. Some are due to concentration on different purposes for evaluation, i.e., accountability versus improvement. Some are due to differences in perspective, e.g., university, state, school district, and accrediting body. Some reflect differences in how evaluation is defined, e.g., should it be competency based or something else? In addition, there were differences in the levels of technical specification; perhaps Borich's models for follow up studies were the most technically prescribed.

In varying degrees, I believe that all of these models are potentially useful for extending the meaning of the definition proposed above. However, I doubt whether any of them is sufficiently comprehensive to ensure that the full intent of the definition would be carried out. In this regard, potential users of the models probably should not become married to any one model. Instead, as Borich and Craig pointed out, evaluators need to be sufficiently informed about the

available alternatives that they can intelligently choose and adapt them for use in particular situations.

In the spirit of proposing alternatives, I will review the particular model of evaluation that I have been developing for the past fifteen years. It is commonly referred to as the CIPP model, which is an acronym drawn from the first letters of the names of the four types of evaluation it proposes--context, input, process, and product evaluation. This model has evolved in relation to practical evaluation experiences, and it reflects an attempt to present a system of concepts by which to apply the definition of evaluation given in the prior section of this paper. In describing this model, I will attempt to relate it to the other models presented in this symposium.

The CIPP model is geared to a systems review of education and human services. In contrast to models such as those proposed by Borich for conducting specific studies, CIPP is geared to providing ongoing evaluation services in an institution. But beyond providing for a dynamic baseline of information such as those proposed by Roth, by Galluzzo, and by the Ohio State group, CIPP also provides for specialized studies designed to help bring about needed improvements.

The orientation towards helping to maintain and improve the quality of institutional operations is illustrated in the flow model which appears in Figure 1. Starting in the left hand corner, it acknowledges that the operations of a teacher education program include various and probably uncoordinated evaluation efforts; ideally these would include the regularized maintenance of a data base such as those described by Galluzzo (he called for seven year cycles of data collection that follow students from admission to performance in the field), Roth, and the Ohio State group. Periodically, however, the flow model indicates that the program needs to undergo a special context evaluation. Such an evaluation would 1) examine current and projected needs for teachers; 2) expose opportunities such as pertinent funding programs, applicable findings from recent research on teaching and teacher education, and school districts and other agencies with a willingness to collaborate with the program; 3) collect and examine perceptions about problems in the program that warrant change; 4) assess the efficacy of institutional goals and priorities; and 5) examine program inputs, processes, and

outputs. Such a context evaluation might be motivated from inside the college or university as a regular "state of the program" assessment or as a response to indications from some sector, e.g., the state's professional education association, of dissatisfaction with the program's services and products. A context evaluation might also be motivated from outside the program, as when NCATE requires a self study or a funding agency requires a "needs assessment" as a basis for justifying a funding request. Such studies may be targeted on specific areas of concern, e.g., the competence of graduates or the research outputs of faculty, or focused more generally on a wide range of program aspects, e.g., students, faculty, curriculum, instruction, research, service, facilities, administration, and policy. In general, such studies aid in system renewal and promotion of better and more efficient service, in diagnosis of particular problems and targeting of improvement efforts, and in communication about the institution's strengths and weaknesses with its constituency and controllers.

The results of the context evaluation, ideally, would lead to a decision about whether to introduce some kind of a change in the system. If decided in the negative (e.g., if given a clean bill of health by NCATE), then the program's staff might well continue with their program operations as usual. However, if a decision to change the program in some way were made, then the program staff would clarify the problem(s) to be solved and formulate their program improvement objectives, and perhaps clarify and validate the teacher competencies they hope to foster. Next, they would consider whether some appropriate program improvement strategy is apparent and readily adaptable to their situation. If so, they would install it and redirect their attentions to using it and evaluating it in the ongoing program.

If no satisfactory solution were apparent, then the staff, according to the flow model, would conduct an input evaluation. Such an evaluation would search the relevant literature, question personnel in other institutions that may have dealt successfully with a similar problem, draw on the ingenuity and creativity of the program's faculty and constituent groups, and possibly would involve outside experts. Subsequently, one or more committees would be assigned to write up one or more proposed improvement strategies. The resulting proposal(s)

would then be assessed against such criteria as responsiveness to the defined needs, problems, objectives, and competencies; theoretical soundness; and feasibility.

The results of the input evaluation would be used to decide whether a sufficiently promising improvement strategy had been found to warrant going ahead with its further development. If not, the faculty would reconsider whether the desired change is sufficiently important to warrant further search, and, if so, would recycle through the search for an improvement strategy. If a promising strategy had been found, then the staff would decide whether or not the strategy could justifiably be installed without further testing. If much were known about the strategy and there was little concern about being able to install it, the faculty would likely turn their attention directly to incorporating the change into regular ongoing activities, without any further specialized evaluation support.

However, if the faculty decided to test it further, they would direct their attention to a controlled examination of the strategy and would subject it to process and product evaluation over whatever period of time would be required to shakedown and debug the procedure and reach the desired level of performance, and readiness for installation. Process evaluation would be employed to assess the degree to which the plan is carried out as intended, while product evaluation would be geared to identifying and assessing the quality of the results.

At some point, if the project has not performed satisfactorily or is viewed as too costly, the leadership of the program might conclude that no further effort is warranted and, in accordance with this conclusion, decide to abort the effort. Such decisions have frequently been made at the conclusion of federally supported projects, when the grantee had to decide whether or not to allocate local funds for the institutionalization of an innovation. As shown in the bottom right hand corner of the flow-chart, even if a project had succeeded, the program's leadership might determine that pertinent conditions had changed sufficiently that the previously desired change was no longer needed, and, accordingly, terminate the effort. Under the assumption that the project was a success and the solution it afforded was still needed and wanted, the faculty would install the proven innovation and

return to normal operations, including regularized evaluation of the ongoing program.

The preceding analysis of evaluation in the context of an institution's change process points up a number of important features of a systems approach to evaluation.

1. Evaluation is seen as an integral part of an institution's regular program and not merely a specialized activity involved in innovative projects, and the implementation of CIPP or any other specialized approach is only a part of the total mosaic of informal and formal evaluation that goes on in an institution.

2. Evaluations have a vital role in stimulating and planning changes.

3. The employment of each type of evaluation in the CIPP Model is indicated only if information beyond what already exists is needed, not by the inherent value in doing each kind of evaluation. In other words, context, input, process, and product evaluations are only a part of a larger milieu of evaluation that goes on in any institution, and the most important function of those commissioned studies is in serving the institution's marginal needs for evaluative information.

4. The development of new programs should include the provision for their ongoing employment and use of evaluation once they have been installed, through something akin to curriculum-embedded evaluation (wherein evaluation is built in to the implementation of a curriculum and, as a matter of course, yields feedback of use in diagnosing, prescribing, and checking progress).

5. Evaluation information not only provides guidance for institutional problem solving, but, if recorded and made available for public review, it also provides a basis for judging whether decisions either to abort or institutionalize a special project were made on defensible grounds.

6. While the CIPP Model makes no special provision for formulating and testing hypotheses about teacher education, it does, through its provision for context, input, and process information provide a rich array of background data against which to interpret and understand outcomes.

The preceding analysis of Figure 1 has shown the CIPP Model in its formative role of motivating and guiding improvement efforts. While I see improvement as the most important purpose of evaluation, I

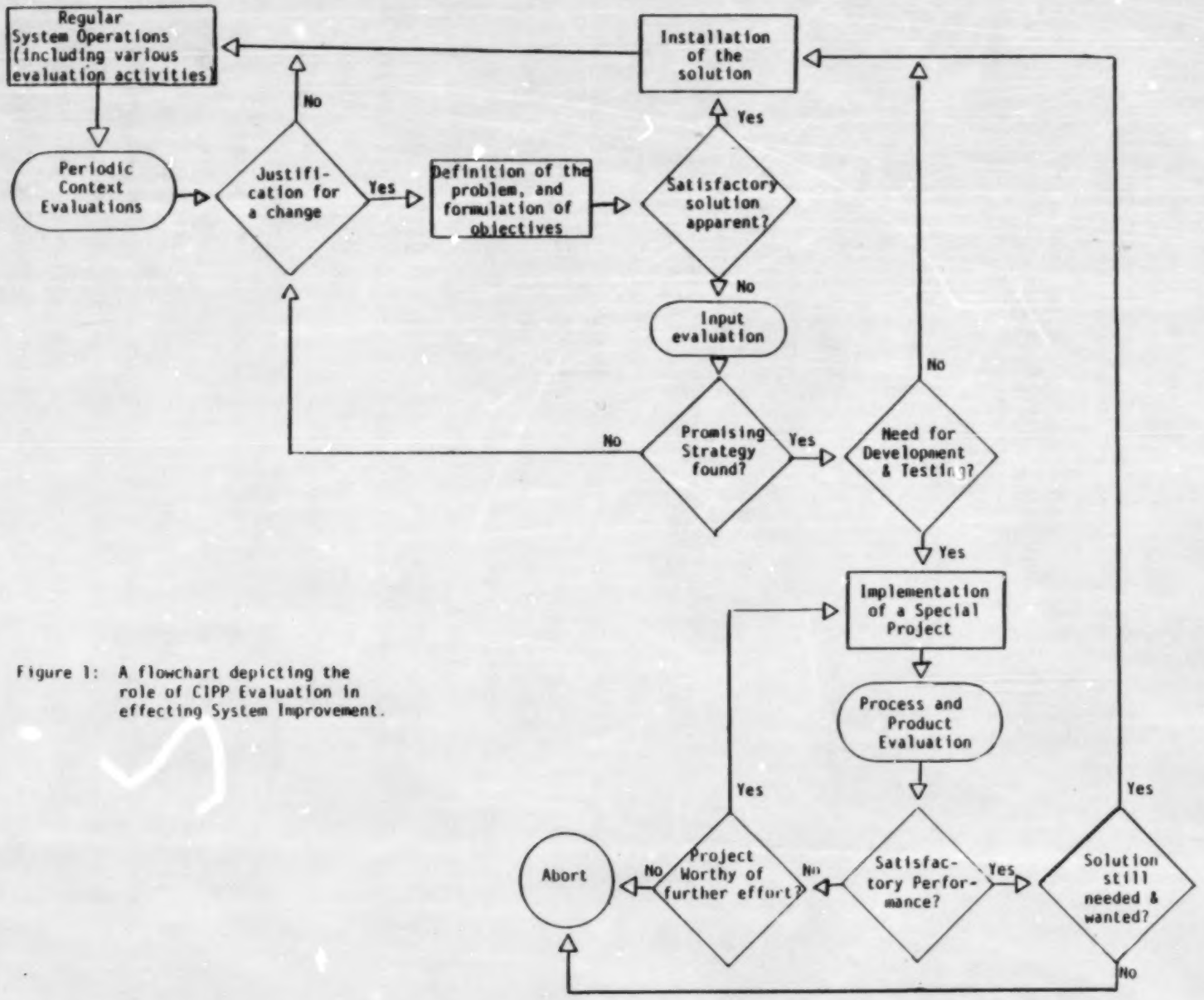


Figure 1: A flowchart depicting the role of CIPP Evaluation in effecting System Improvement.

acknowledge that teacher educators must also be accountable for the quality of their work, and I believe that CIPP can make an important contribution as well.

As shown in Table 2, context, input, process, and product evaluations may be used both to guide improvement (the formative role), and to supply information for accountability (the summative role). While this chart shows that evaluators would design and conduct evaluation so as to assist a faculty to plan and implement their program, it also suggests that regardless of the narrowness or breadth of the information requirements of program developers, the evaluators should also keep in mind and try to address the full range of information needs of external audiences that someway would want to form conclusions about the improvement effort. Moreover, they would maintain a record of the information collected and evidence of the extent that the developers used it to guide their work. While such information would not answer all the questions of an external summative evaluation group, such as NCATE, it would certainly help in answering some of them. Especially a full implementation of the CIPP approach would yield information of use in addressing the following questions:

1. What needs (in society at large, in schools, and of teacher trainees) were identified, how pervasive and important were they, and to what extent were the program's goals, objectives, and competency statements reflective of assessed needs (addressed by context evaluation)?

2. What program plan was adopted to address the needs, what alternatives were considered, why was it chosen over them, and to what extent was it a reasonable, potentially successful, and cost effective proposal for meeting the assessed needs (addressed by input information)?

3. To what extent was the program plan implemented, and how and for what reasons did it have to be modified (addressed by process information)?

4. What results--positive and negative as well as intended and unintended--were observed, how did the various stakeholders judge the worth and merit of the outcomes, and to what extent were the needs of the target population met (product information)?

	EVALUATION TYPES			
	CONTEXT	INPUT	PROCESS	PRODUCT
DECISION MAKING (formative orientation)	Guidance for choice of objectives and assignment of priorities	Guidance for choice of program strategy Input for specification of procedural design	Guidance for implementation	Guidance for termination, continuation, modification, or installation
ACCOUNTABILITY (summative orientation)	Record of objectives and bases for their choice along with a record of needs, opportunities, and problems	Record of chosen strategy and design and reasons for their choice over other alternatives	Record of the actual process	Record of attainments and recycling decisions

Table 2: The Relevance of Four Evaluation Types to Decision Making and Accountability

3. How should the CIPP Model be operationalized in the context of improving teacher education programs?

If the CIPP Model is to be of use to teacher educators, they will need to convert its concepts into actions. I'll try to help move us in this direction by explicating context, input, process, and product evaluation and by citing techniques or processes that I see as particularly useful in each of these types of evaluation. Then I will discuss the general problems involved in designing any particular study.

Table 3 has been provided to assist in explicating each of the four types of evaluation. It defines context, input, process, and product evaluation in relation to their objectives, methods, and uses.

Context Evaluation

As I have already indicated, the primary orientation of a context evaluation is to identify the strengths and weaknesses of a program, and to provide direction for improvement. The main objectives of this type of study are to assess the program's overall status, to identify its deficiencies, to inventory the strengths at hand that could be used to remedy the deficiencies, to diagnose problems whose solution would improve the program and, in general, to characterize the program's environment. Incidentally, the Erly paper includes an excellent portrayal of the current difficult context of teacher education in general. A context evaluation is also aimed at examining whether existing goals, priorities, and defined competencies are attuned to the needs of the schools being served, the students being taught, and the teacher education faculty who have a continuing need for staff development. Whatever the focal program (or part of a program), the results of a context evaluation should provide a sound basis for adjusting (or establishing) goals, priorities, and competency statements, and for targeting needed changes.

The methodology of a context evaluation may involve a variety of measurements of the program of interest and various types of analysis. A usual starting point is to interview the clients of the study (e.g., university administrators and faculty members) in order to obtain their perceptions of strengths, weaknesses, and problems. Hearings, community

	CONTEXT EVALUATION	INPUT EVALUATION	PROCESS EVALUATION	PRODUCT EVALUATION
OBJECTIVE	To define the institutional context, to identify the target population & assess their needs, to identify opportunities for addressing the needs, to diagnose problems underlying the needs & to judge whether proposed objectives are sufficiently responsive to the assessed needs.	To identify & assess system capabilities, alternative program strategies, procedural designs for implementing the strategies, budgets, & schedules.	To identify or predict, in process, defects in the procedural design or its implementation, to provide information for the preprogrammed decisions, and to record & judge procedural events & activities.	To collect descriptions & judgments of outcomes & to relate them to objectives & to context input, & process information, & to interpret their worth & merit.
METHOD	By using such methods as system analysis, survey, document review, hearings, interviews, diagnostic tests, & the Delphi technique.	By inventorying & analyzing available human & material resources, solution strategies, & procedural designs for relevance, feasibility & economy. And by using such methods as literature search, visits to exemplary programs, advocate teams, & pilot trials.	By monitoring the activity's potential procedural barriers & remaining alert to unanticipated ones, by obtaining specified information for programmed decisions, by describing the actual process, & by continually interacting with, & observing the activities of project staff.	By defining operational & measuring outcome criteria, by collecting judgements of outcomes from stakeholders, & by performing both qualitative & quantitative analyses.
RELATION TO DECISION-MAKING IN THE CHANGE PROCESS	For deciding upon the setting to be served, the goals associated with meeting needs or using opportunities, & the objectives associated with solving problems, i.e., for planning needed changes. And to provide a basis for judging outcomes.	For selecting sources of support, solution strategies, & procedural designs, i.e., for structuring change activities. And to provide a basis for judging implementation.	For implementing and refining the program design and procedure, i.e., for effecting process control. And to provide a log of the actual process for later use in interpreting outcomes.	For deciding to continue, terminate, modify, or refocus a change activity. And to present a clear record of effects (intended & unintended, positive & negative).

TABLE 3: Four Types of Evaluation

forums, and further interviews may be conducted to generate additional hypotheses about what changes are needed. These hypotheses may be used to construct a survey instrument and it may be administered both to a carefully defined sample of stakeholders (e.g., students, school district administrators, professors, university placement officials, and state certification officials) and made available more generally to anyone who wishes to provide input, with the analyses of the two sets of responses kept separate. Existing records (e.g., alumni files, student grades and test scores, and correspondence from employees) should also be examined to identify performance patterns and background information; this recommendation is especially applicable to institutions that maintain data bases such as those recommended by Roth and Zimpher. Special diagnostic tests might be devised and administered, especially in the case of competency-based programs. The papers by Galluzzo and Erly suggest competency clusters that might be used to construct such tests. An expert review panel might be engaged to visit, closely observe, and judge the worth and merit of the program; of course Gardner's paper is of use here. Throughout the study, an advisory committee, representative of the various stakeholder groups, might be involved in clarifying the evaluative questions and interpreting the findings; such a group might assist in overcoming the problem of faculty acceptance of evaluative findings identified by Sandefur. A consensus-building technique, such as Delphi, might be used to secure agreements among faculty and administrators about priority needs. A workshop might be conducted to help the clients of the evaluation to study and apply the findings. The outcome of such a workshop might be a five year plan such as that projected by Erly; you may recall his suggestion that such a plan should: 1) project system needs; 2) project curriculum development changes and associated staff development needs; 3) identify student performance deficiencies; 4) project mandated program requirements and associated staff development needs; and 5) identify what teachers see as their needs.

A context evaluation may have a number of constructive uses. It may provide a means by which a teacher education faculty communicates with its parent institution to gain a shared conception of the program's strengths and weaknesses, needs and opportunities, and priority

problems. It may be used to convince a funding agency that a proposed project is directed at an area of urgent need or to convince the parent institution's central administration that a proposed budget is justified. It might be used to formulate objectives for staff development and/or curriculum revision. It could be used to select particular program components for priority assistance. Of course it should often be used to help students and their advisors to focus their attention on developmental areas where more progress is needed. Also, it could be used to help decide how to cut programs, hopefully to help a teacher education program get stronger while getting smaller. Periodically, such a study would be used to help a program justify its application for accreditation and, when indicated, to target the areas where improvements are needed.

I see these as the major instances of how a context evaluation could assist individuals and groups to set priorities for improvement efforts. Another use may come later if there is a need to assess and demonstrate what has been accomplished through an improvement project. This is especially the case when a program, having previously been denied accreditation, reapplies. One basis for judging the outcomes of a program improvement effort is by assessing whether they are adequately responsive to the needs that were identified through the previous context evaluation. In general, context evaluation records are a pertinent means by which the faculty of a teacher education program can defend and/or adjust the efficacy of their goals and priorities.

Input Evaluation

The main orientation of an input evaluation is to help prescribe a program by which to bring about needed changes. It does this by searching out and critically examining potentially relevant approaches. Unless I missed it at the symposium, discussion of this crucial type of evaluation was lacking and, even though Galluzzo's model ostensibly is based on the CIPP Model, his paper essentially ommitted input evaluation. In my view, an input evaluation is a crucial precursor of the success, failure, and efficiency of a change effort. Change projects are constrained by initial decisions about how resources will

be allocated, and a potentially effective solution to a problem will have no possibility of impact if a teacher education faculty does not at least identify it and assess its merits when they are planning their change project.

Essentially, an input evaluation should identify and rate relevant approaches (including the one[s] already operating in the main program of interest) and assist in explicating and "shaking down" the one that is chosen for installation or continuation. It should also search the environment of the teacher education program for barriers, constraints, and potentially available resources that ought to be considered in the process of activating the program. The overall intent of an input evaluation, in the context of teacher education, is to help the faculty and administration consider alternative program strategies in the context of their needs and environmental circumstances and to evolve a plan that will work for them; another important function is to help them avoid the wasteful practice of pursuing proposed innovations that predictably would fail or at least waste resources. Sample inputs, i.e., program elements, include increased field experiences, extended student teaching, more hours in general studies course work, federal or state mandated changes in curriculum (e.g., P.L. 94-142), mainstreaming and multicultural education.

The methods involved may be described in a series of stages, although there is no set sequence of steps for conducting an input evaluation. One might begin by reviewing the state of practice with respect to meeting the specified needs; this could be done by reviewing relevant literature, visiting exemplary programs, consulting experts and representatives of the state department of education, querying pertinent information services, and, of course, inviting proposals from the involved faculty. This information might be organized in a special planning room and subjected to indepth investigation by a special study group. Their investigation might be conducted over a period of time through a special decision seminar. This group might use the information to assess whether potentially acceptable solution strategies exist. In addition, they might rate promising approaches for potential effectiveness and feasibility, and advise their fellow faculty members and their administrators about whether to seek a novel solution. If an

innovation is to be sought, the seminar group might assist the faculty and administration to define criteria to be met by the innovation, structure a request for proposal, obtain competing proposals, and rate them for potential effectiveness and feasibility. Subsequently, the study group might analyze and rank the potentially acceptable proposals and suggest how their best features could be combined. In addition, the team might conduct a type of hearing in which faculty members and administrators are invited to express concerns and to give their realistic appraisal of the resource problems, political barriers, etc. that need to be dealt with in the process of installing the solution. Finally, the seminar group might prepare a plan, laying out both short-range and long-range recommendations, for implementation of the proposed strategy.

The Advocacy Team Technique is a relatively new procedure for conducting input evaluation that deserves special mention. This technique is especially applicable in situations where appropriate program strategies that have been demonstrated to help meet needs similar to those in question are not available. Two or more teams of persons with applicable expertise and experience are convened; they are given the context evaluation results and the objectives for which a program is needed, provided specifications for designing a program proposal, and oriented to the criteria by which the competing responses will be judged. The teams competitively prepare their proposals, but in isolation from each other. Their reports are rated by an appropriately qualified panel and/or pilot tested in accordance with pre-established criteria. Subsequent steps involve members of the user system in operationalizing the winning strategy or combining and operationalizing the best features of the two or more competing strategies. Advantages of the Advocacy Team Technique are that it provides 1) an explicit procedure for generating and assessing competing program strategies; 2) an explicit accountability record of why a particular solution strategy was selected; 3) a forum that exploits bias and competition in a constructive search for alternatives; and 4) a means of involving personnel from the adopting system, either as advocacy team members, or as members of the team that performs the convergence and operationalization of activities following the ranking of the competing strategies.

Additonal information, including a technical manual and results of five field tests of the technique, is available in a doctoral dissertation by Diane Reinhard (1972).

Input evaluations have a number of applications. A chief one is in preparing a proposal for submission to a funding agency or a program's policy board. Another is to assess one's existing program--whether or not it seems to be working--against what is being done elsewhere and proposed in the literature. Input evaluations may be used to prepare plans for overcoming deficiencies reported by accrediting teams. Another use is to provide a structure and forum by which historically antagonistic groups can reach agreement on some course of action. In addition, the records from an input evaluation study help decision makers to be accountable for their choice of one course of action above other possibilities.

Process Evaluation

In essence, a process evaluation is an ongoing check on the implementation of a plan. One objective is to provide feedback to managers and staff about the extent to which the program activities are being carried out as planned, are using the available resources in an efficient manner, and are on schedule. Another is to provide guidance for modifying or explicating the plan as needed, since not all aspects of a plan can be determined in advance, and since some of the initial decisions may later prove to be flawed. Still another objective is periodically to assess the extent to which program participants accept and are able to carry out their roles. Finally, a process evaluation should provide an extensive record of the progam that was actually implemented, how it compared to what was intended, a full account of the various costs incurred in carrying it out, and overall how observers and participants judged the quality of the effort.

The key to a sound process evaluation is the process evaluator. More often than not, a program staff's failure to obtain guidance for implementation and to document their activities is due to a failure to assign anyone to do this work. Erroneously, it is too often assumed that the department chairperson and faculty can and will do an adequate

job of process evaluation as a normal part of their assignments. While some review and documentation can be done through routine activities, such as faculty meetings, these are not a sufficient means of meeting the requirements of a sound process evaluation. In my experience, these requirements can be met well only by assigning one or more persons to provide ongoing review, feedback, and documentation.

The process evaluator will find that there is much work to be done. The following scenario is provided as an illustration of what he might do. Initially, the process evaluator could review the program plan and any prior evaluation on which it is based to identify important aspects of the program that should be monitored. Some examples that might be identified are faculty development workshops, recruitment materials and procedures, proficiency testing, advising of students, tutoring services, faculty planning, classroom instruction, field experience, and placement services. Galluzzo, in his discussion of process evaluation, recommended especially that students and faculty be asked to characterize their experiences in the program, to help identify the reasons why some students drop out or transfer, and to offer judgments about the value of the program offerings and the quality of instruction. As another means of identifying what should be looked at the evaluator might form an advisory group, broadly representative of program participants, and periodically ask them to identify concerns and questions that should be addressed. Other questions of relevance will occur to the evaluator as he observes program activities.

With questions and concerns in mind, such as those mentioned above, the process evaluator could develop a general schedule of data collection activities and begin carrying them out. Initially, these should probably be as unobtrusive as possible in order not to threaten program staff or get in their way. Subsequently, as rapport is developed, the process evaluator can use a more structured approach. At the outset, the process evaluator might try to get an overview of how the program is operating by visiting and observing centers of activity, reviewing program documents, attending faculty meetings, and interviewing key participants. He then could prepare a brief report that summarizes his data collection plan, reviews what he has learned, and points out what he sees as key issues. He then could present this

report at a faculty meeting and invite the faculty chairperson to lead a discussion of it and to use it for program revision as he and his faculty see fit. Later in the meeting, the process evaluator could review with the faculty his plans for further data collection and a subsequent report. He could ask for their reactions about what feedback would be most useful at a subsequent meeting, also their suggestions about how to best obtain certain items of information, e.g., observations, faculty kept diaries, interviews, or questionnaires. On the basis of feedback from the faculty, the evaluator would schedule future feedback sessions, modify the data collection plan as appropriate, and proceed accordingly. He should continually demonstrate that the main purpose of process evaluation is to assist the faculty in carrying out their program, through a kind of quality assurance process. Throughout this interactive process, the evaluator should periodically prepare and file reports on his perception of the extent that the program plan has been implemented. He should describe main deviations from the plan, and should make special note of variation within the program concerning how different persons and subgroups are carrying out the plan. He should also characterize the ongoing planning activity and trace the evolution of the basic plan on which the program is based.

The main use of process evaluation is to obtain feedback that can aid staff to carry out a program as it was planned, or, if the plan is found to be seriously flawed, to modify it as needed. Some managers see regularly scheduled process evaluation feedback sessions as a means of keeping staff "on their toes" and abreast of their responsibilities. Process evaluation records are also useful for accountability, since funding agencies, policy boards, accrediting agencies, and constituents typically want to know whether grantees did what they proposed or had been charged to do. Process evaluations can also help external audiences to learn what was done in the program in case they want to conduct a similar one. And a process evaluation is a vital source of information for interpreting product evaluation results, since in considering why program outcomes turned out as they did one would want to know what was actually done in carrying out the program; in this respect, process evaluation, in addition to promoting improvement and

supporting accountability, also fosters understanding of teacher education.

Product Evaluation

The purpose of a product evaluation is to measure, interpret, and judge the attainments of a program. Feedback about what is being achieved is important both during a program cycle and at its conclusion. Also, as illustrated by Borich's paper on follow up studies, product evaluation often should be extended to assess long-term effects. The main objective of a product evaluation is to ascertain the extent to which the program has helped to meet the needs of the groups it is intended to serve, e.g., teacher trainees, schools, and the society at large. In addition, a product evaluation should look broadly at the effects of the program, including intended and unintended effects and positive and negative outcomes. A product evaluation should gather and analyze judgments of the program's success from a broad range of people associated with the program, e.g., students, instructors, administrators, employers, accreditation officials, and visiting experts. Sometimes it should compare the outcomes of the program under study with those of alternative programs. Frequently, the client wants to know how the attainments compare to previously stated objectives or designed competencies. Also, usually it is quite important to offer interpretations of the extent that failure to achieve objectives, meet needs, or satisfy competencies was correlated with a failure to implement the program plan; here again we see the importance of process evaluation records. In today's difficult economy, clients are also acutely concerned about the extent to which the outcomes are worth more than the cost of attaining them. Finally, a product evaluation usually should view outcomes from several vantage points: in the aggregate, by subgroupings of program recipients who might be differentiated by needs and services received, and sometimes by individuals. An outcome associated with an individual may be classified as a success or failure depending on whether it has satisfied a diagnosed need of the individual; such product evaluation at the level of individuals also allows aggregation across individuals to get an overall index of the

extent to which the program has succeeded in meeting the collective and differential needs of individuals.

As illustrated by the range of product evaluation-oriented models presented at the symposium there is no set algorithm for conducting a product evaluation, but there are many applicable methods. The following scenario is intended to illustrate the range of techniques that might be employed. In general, a combination of techniques should be used to obtain a comprehensive view of effects and to provide crosschecks on the various findings.

The product evaluators might begin by assessing performance in relation to some previously chosen standard. Such assessments might be based on test performance compared to a profile of previously assessed needs, pretest performance, selected norms, specified competencies, or the performance of a comparison group. The tests used might be published objective tests, specially made criterion referenced tests, or applied performance tests (see Sanders, 1977). Performance assessments might also be based on ratings of performance by observers, instructors, employers, and/or program recipients themselves. And, experts might assess work products and compare them to previously developed needs profiles for the program recipients who produced them.

Borich provided focused, technical advice for conducting three types of studies to follow up graduates. Each is based on prestated competencies, and specific steps are listed. These are excellent as technical recommendations to evaluators of competency-based programs, but, in isolation each would yield a much too narrow range of information to constitute an adequate product evaluation.

In order to assess performance beyond that related to intended outcomes, evaluators need to make an extensive search for unanticipated outcomes, both positive and negative. They might conduct hearings or group interviews to generate hypotheses about the full range of outcomes, and follow these up with clinical investigations intended to confirm or disconfirm the hypotheses. They might conduct case studies of the experiences of a carefully selected sample of participants in order to obtain an indepth view of the effects of the program. They might survey, via telephone or mail, a sample of participants to submit concrete examples, e.g., pieces they have written of how the project has

influenced their work. They might engage observers to view the performance of program and comparison groups in regular settings and to develop and validate tests that distinguish between their performance and thus give a view of the unique contributions of the program, pro and con (see Brickell, 1976). They might search out and examine program outcomes in relation to a comprehensive checklist of outcomes that have been observed for similar programs. As a final example, they might conduct a "jury trial" by which to introduce and examine all available evidence that reflects on the success or failure of the program (see Wolf, 1974).

Reporting of the product evaluation findings may occur at different stages. Interim reports may be submitted during each program cycle to indicate the extent the targeted needs are being addressed and met. End-of-cycle reports may sum up the results achieved and interpret them in the light of preassessed needs, costs incurred, and the extent the plan was carried out. Follow up reports may also be submitted to indicate what if any long-term impacts can be found. In such reports, the results might be analyzed in the aggregate, for subgroups, and/or individuals.

The basic use of a product evaluation is to determine whether a given program is worth continuing, repeating, and/or extending into other settings. It also should provide direction for modifying the program so that it better serves the needs of all members of the target audience and so that it will become more cost effective. Of course, it should help potential adopters of the program to decide whether it merits their serious consideration. Product evaluations have psychological implications, since by showing signs of growth and/or superiority to competing efforts they reinforce the efforts of both staff and program recipients; likewise they may dampen enthusiasm when the results are poor. Product evaluation information is an essential component of an accountability report, and when there is evidence of significant achievement it can aid in securing additional financial and political support from the community and funding agencies. When this information reveals no important gains that are warranted by the associated costs, product evaluation can help avoid continued wasteful investments in the program. Moreover, a record of the results obtained,

especially in consideration of the program approach used and the costs involved, can assist other developers to decide on the wisdom of pursuing a similar course of action.

The preceding discussion indicates that context, input, process, and product evaluation serve unique functions, but that a symbiotic relationship exists among them. It further shows that a variety of methods are applicable to each type of evaluation. But it doesn't deal with the evaluator's practical problem in deciding which methods to employ in a particular study.

Designing Evaluations

To guide the implementation of an evaluation--whether context, input, process, or product evaluation (or some combination)--teacher educators obviously need to design the work to be done. This involves preparing the preliminary plans and subsequently modifying and explicating them as the evaluation proceeds. These plans must deal with a wide range of choices pertaining to the conduct of the evaluation, e.g., the key audiences and questions; the program or project to be assessed; whether a context, input, process, and/or product evaluation is indicated; the timing and location of the evaluation; the extent and nature of controls to be imposed; the contrasts to be made; the sources of needed information; the methods; instruments and schedule for data collection; the formats and procedures for labeling, storing, and retrieving information; the methods of analysis and interpretation; provisions for communicating findings; and criteria and arrangements for assessing the evaluation results. Decisions about such evaluation activities form the basis for contracting and financing the evaluation work, working out protocol with the involved institutions, staffing the evaluation, and scheduling and guiding staff activities.

We might wish that evaluators could finalize design decisions at the outset, and then follow them precisely. However, the dynamic and interactive qualities of many evaluations, plus their service orientation make difficult, if not impossible, the accurate, long-range projection of specific information needs. Consequently, technical plans for data collection and analysis, made prior to the start of a study,

often are based on erroneous assumptions and found later to be inappropriate or incomplete. Rigid adherence to the original evaluation design--especially if it had been defined in specific terms--often would detract greatly from the utility of the study by directing it to the wrong questions, basing it on an inappropriate set of assumptions, and/or convincing members of the audience that the evaluator has an ivory tower orientation.

Hence, those charged with planning evaluations are faced with a dilemma. On the one hand they need to carefully plan their evaluation activities so that they can carry them out efficiently and with an acceptable amount of rigor, and convince their clients that they know what they are doing. On the other hand, they need to approach the design of evaluation studies flexibly and provide for periodically reviewing and otherwise modifying the design so that the evaluation remains responsive to the needs of the audiences. This dilemma is especially troublesome to evaluators, since clients often expect or demand up-front technical designs and later become disenchanted with rigid adherence to the original design if it yields much information that is no longer perceived as useful. The client often perceives that somehow the evaluator should have been smarter in projecting information needs, and more skilled in planning the data collection activities.

To address this dilemma evaluators must view design as a process, not a product; and they need to get their clients to do likewise. Evaluation goals and procedures should be sketched in advanced, but periodically they should be reviewed, revised, expanded, and operationalized. Fundamentally, this process should be guided by a defensible view of what constitutes sound evaluation, by a sensitivity to factors in the real world which often interfere with evaluation work, and by ongoing communication, about the pertinence and adequacy of the design, between the evaluators and their audiences.

At the outset of the process, I believe it is important to listen and probe. Who are the primary clients? What do they want from the evaluation? Why? What type(s) of evaluation (context, input, process, product) would be most responsive? How do the clients think the evaluation should be conducted? What time line do they have in mind? Who do they see as the main audience? Who might "get hurt" as a

consequence of the evaluation? Why? Whose cooperation will be essential? What information already exists? What's the relevant history? Realistically, what positive benefits from the evaluation could be expected? What deleterious effects are real possibilities, and how could they be avoided? What qualifications are required to do the job?, etc. Whenever there is a choice, those considering doing an evaluation should pursue questions like these before agreeing that an evaluation should be done or that they are the right persons to do it.

Assuming a positive decision to go ahead, the designated evaluator should sketch an overall plan. This plan should take into account what the evaluator has learned about the setting and particular needs for the evaluation, and it should conform to generally accepted standards of sound evaluation. In addition, it should speak, at least in a general way, to the full range of tasks to be done.

Table 4 provides a general outline of the points to be addressed in an evaluation design. These points are applicable when developing the initial design or later when revising or explicating it. Of course, they serve only as general indications of the detailed information that eventually must be provided to flesh out and operationalize the design.

The formulation of the design requires that the client and those engaged to do the evaluation collaborate, from the outset, when they must agree on a charge. The client needs to identify the object, e.g., the program to be evaluated, and the evaluator can help by guiding the client to define clear and realistic boundaries around what will be looked at. The client is a prime source for identifying the various groups with potential interest in the study; but the evaluator also needs to touch base with the potential audiences and think about the evaluation within the relevant social context in order to identify the full range of legitimate audiences. The client and other audiences need to identify the purpose of the study, i.e., indicate what information they need and how they plan to use it, and the evaluator needs to pursue clarifying questions in order to sort out different (perhaps conflicting) purposes and to get the client to assign priorities. The evaluator needs to indicate what general type(s) of study (e.g., context, input, process, and/or product) seems needed, and the client should confirm this general choice or help to modify it. In rounding

TABLE 4

Outline for Documenting Evaluation Designs

Review of the Charge

- Definition of the object of the evaluation
- Identification of the client and audiences
- Purpose(s) of the evaluation
- Type of evaluation (e.g., context, input, process, or product) to be employed
- Principles of sound evaluation (i.e., standards) to be observed

Plan for Obtaining Information

- The general strategy (e.g., survey, case study, field experiment, or site visitation)
- Working assumptions to guide measurement, analysis, and interpretation
- Collection of information (i.e., sampling, instrumentation, and data collection)
- Organization of information (i.e., coding, filing, and retrieving)
- Analysis of information (both qualitative and quantitative)
- Interpretation of findings (e.g., against needs, objectives, values, or other programs)

Plan for Reporting the Results

- Preparation of reports
- Dissemination of reports
- Provision for follow-up activities to promote impact of the evaluation

Plan for Administering the Study

- Summarization of the evaluation schedule
- Plan for meeting staff and resource requirements
- Provision for metaevaluation
- Provision for periodic updating of the evaluation design
- Budget
- Memorandum of agreement or contract

out the charge, the evaluator needs to make clear that the evaluation will be conditioned to meet a certain set of standards, and the client should be asked to help select and assign priorities to the applicable standards.

Basically, the plan for obtaining information should be worked out by the evaluator, but it should be subjected to careful review by the client and modified accordingly. The evaluator should provide an overview of the general strategy to be employed (e.g., survey, case study, site visitation, advocacy teams, goal-free search for effects, adversary hearings, or field experiment), and technical plans for collecting, organizing, and analyzing the needed information. While the clients should at least react to the technical plans, they should exert a major influence in deciding how the findings will be interpreted (e.g., against objectives, against the results of prior needs assessments, based on the evaluator's professional judgment, or through some type of formal group process). The evaluator and client should anticipate that the plan for obtaining information will likely change and expand during the course of the evaluation, as new audiences are identified and information requirements change.

The part of the evaluation design devoted to reporting results should be geared to promote utilization. The client and audience should be involved in projecting the contents and timing of needed reports. They should also assist in planning how the results will be disseminated (i.e., organized, displayed, delivered, reviewed, revised, and documented for later use). Moreover, the client and evaluator should seriously consider whether the evaluator might play an important role--beyond the delivery of the final report--in helping the client and audience to apply the findings to their work. Overall, the plan for reporting should be directed to promote impact through whatever means seem appropriate (e.g., oral reports and hearings, multiple reports targeted to specified audiences, press releases, socio-dramas to portray and explore the findings, and workshops aimed at applying the findings).

The final part of the design, the plan for administering the study, is oriented towards operationalizing the conceptual and technical plans. The evaluator needs to identify and schedule the evaluation tasks, consistent with the client's needs and in consideration of the relevant

practical constraints. Staff who carry out the evaluation work and special resources (such as office space and data processing facilities) need to be identified; and the client needs to assure that the proposed personnel have the necessary level of credibility as viewed by the audiences. The evaluator and client need to agree on how the evaluation plans, processes, and reports will be assessed against the agreed upon standards. They also should agree on a mechanism by which periodically to review, update, and document the evolving evaluation design. They need to lay out a realistic budget. And, in my view, they should summarize and formalize their general agreements about the form and function of the evaluation in a memorandum of agreement or a contract.

The foregoing discussion of Table 4 has been necessarily general; but it indicates that designing an evaluation is a complex and ongoing task. It recommends continued collaboration between evaluator, client, and other audiences; and it emphasizes the importance of evolving the evaluation design in order to serve emerging information requirements. Also, it emphasizes the need to maintain professional integrity in the evaluation work.

Conclusion

The Bible illustrates the hazzards of affluence by stating that it is easier for a rich man to pass through the eye of a needle than to enter heaven. Likewise, it may be easier for professors to give up their academic freedom than willingly submit their programs to evaluation. Or, to paraphrase Edgar Dale, all one has to do to evaluate a higher education program is to convince every involved professor to withhold his or her pocket veto. Surely, anyone who has ever tried to evaluate a higher education program can attest to the difficulties involved.

But the symposium for which this paper was written is testimony that progress is being made, albeit slow. The working papers presented at the meetings evidence careful thinking about the problems of evaluating teacher education, as well as some concrete field experiences. Moreover, the ensuing discussions evidenced a commitment from many sectors for improving evaluations of teacher education. What

appears to be a heightened interest in evaluating teacher education, perhaps, may best be understood in terms of this field's current troubled context. I think Maurice Erly made this relationship very clear. It is always easier to evaluate when one's future otherwise appears to be in jeopardy.

For those who are committed to the position that evaluation is a fundamentally important concomitant of improvement, this is a good time aggressively to promote sound evaluation work. The field has a readiness to accept anything that will foster quality, efficiency, and public credibility. And, I believe that sound evaluation potentially offers substantial assistance here. I have, therefore, welcomed the opportunity to explore with leaders in the field of teacher education possible ways of advancing the practice of "teacher education evaluation."

This paper is based on the position that the most important purpose of evaluation is not to prove but to improve. We cannot make our teacher education programs work better unless we know where they are weak and strong and unless we become aware of better approaches. We cannot be sure that our goals, priorities, and defined competencies are worthy unless we can match them to the needs of the students and schools they are intended to serve. We cannot plan effectively if we are unaware of options and their relative merits. And, we cannot convince our constituents that we have done good work and deserve continued support unless we can show them evidence that we have done what we promised and have produced beneficial results. For these and other reasons, teacher educators must subject their work to competent evaluation. It must help them sort out the good from the bad, point the way to needed improvements, be accountable to their sponsors and clients, and, in general, promote a better understanding of teaching and teacher education.

This paper has placed highest priority on the improvement function of teacher education evaluation. It has proposed a detailed definition of evaluation and extended this definition to show how evaluation can be used to guide improvement efforts and serve accountability needs. The core concepts of context, input, process, and product evaluation were defined in some detail and related to teacher education. Finally,

general guidelines for designing evaluation studies were provided. Overall, this paper is a partial response to three of the important questions about evaluation of teacher education that were raised in the symposium. I hope the paper will help stimulate further discussion and increased efforts to improve teacher education through evaluation. I will welcome feedback from such efforts.

References

Borich, G. D. Three models for conducting follow-up studies of teacher education and training. A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Borich, G. D. Evaluation models: A question of purpose not terminology. Unpublished manuscript. Austin: University of Texas, 1982.

Brickell, H. M. Needed: Instruments as good as our eyes. Occasional Paper Series, No. 7. Kalamazoo, MI: Evaluation Center, Western Michigan University, July, 1976.

Craig, S. B., Jr. Experimental program evaluation. A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Cronbach, L. J. and Associates. Toward reform of program evaluation. San Francisco: Jossey-Bass Publishers, 1980.

deVoss, G., Lemish, P., & Zimpher, N. L. A system for documenting and evaluating the experiences of pre/inservice teachers. The College of Eduation, Ohio State University, 1981. (unpublished)

Erly, M. C. A practitioner's perceptions regarding problems in assessing the effectiveness of teacher education programs. A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Galluzzo, G. R. Program evaluation in teacher education: From admissions to follow-up. A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Gardner, W. E. NCATE accreditation: Problems, issues, and needed research. A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Guba, E. & Lincoln, Y. The distinction between merit and worth in evaluation. Paper presented at the annual meeting of the Evaluation Network, Cincinnati, Ohio, September, 1979.

Patton, M. Q. Utilization-focused evaluation. Beverly Hills/London: Sage Publications, 1978.

Reinhard, D. L. Methodology development for input evaluation using advocate and design teams. Unpublished doctoral dissertation, Ohio State University, 1972.

Roth, R. A. Requirements of a data base for effective program evaluation. A working paper of the symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Sandefur, J. T. Teacher education's evaluation of graduates: Where are we going and how do we know when we get there? A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Sanders, J. R. & Sachse, T. P. Applied performance testing in the classroom. Journal of Research and Development in Education, Spring, 1977, 10, 92-104.

Scriven, M. Evaluation thesaurus. Point Keyes, CA: Edge Press Publishing, 1980.

Stake, R. E. The countenance of educational evaluation. Teachers College Record, 68(7), April, 1967.

Wolf, R. L. The application of select legal concepts to educational evaluation. Unpublished doctoral dissertation, University of Illinois, 1974.

PROGRAM EVALUATION IN TEACHER EDUCATION:

FUTURE DIRECTIONS

Marlene I. Strathe

Educational Research and Development Center
University of Northern Iowa

PROGRAM EVALUATION IN TEACHER EDUCATION:

FUTURE DIRECTIONS

Marlene I. Strathe
Educational Research and Development Center
University of Northern Iowa

Proposing future directions for program evaluation efforts in teacher education or, at the very least, speculating about those directions may be analogous to suggesting where, how, and when a group of interested fishermen might go fishing. The possible membership in such a fishing group ranges from those interested in catching sunfish to hearty souls determined to land a swordfish. It may even be questioned whether all agree on the purpose of the trip; is it actually to catch fish or simply to enjoy the process of fishing? The discussion and planning of such a trip becomes even increasingly more complex when "expert fishermen," reflecting a variety of interests, attempt to discuss where they should fish, the appropriate bait and tackle, equipment needed, and how the success of the fishing trip might be judged.

Comparing this analogy to the state of program evaluation in teacher education reveals some striking similarities. Numerous "expert fishermen" have been discussing the "trip" for quite some time; and, while general agreement appears to exist regarding the value of such a venture, beyond this point agreement dissipates rapidly. The fundamental question of the purpose for evaluation, i.e., accountability, decision-facilitation, or both, remains a point of debate. Beyond this, differences continue to exist as to which model, instruments, approaches, audiences, and so forth are of major concern and focus. Thus, any attempt to delineate future directions must reflect the recognition that differences continue to exist, but that some common threads have emerged from previous efforts. An examination of these common elements appears to be in order before postulating on future directions.

Elements of Agreement

First, without question, those who have been involved and interested in program evaluation efforts believe that these efforts are both valuable and necessary. In addition, most believe that these efforts should have some impact on the design and delivery of teacher preparation programs. Roth (1982) stated "the most important reason [for conducting program evaluation] should be in order to improve programs and the quality of its graduates." Galluzzo (1982) concurred when he stated the role of program evaluation should be "to create and stimulate critical discussion about the goals and objectives of a teacher education program." But, despite the perceived need for evaluation and the belief that evaluation should have an impact on programs, one wonders if this premise has been internalized by any others outside of those who are, by virtue of their position, directly interested in or responsible for evaluation. Thus, though we may be enthusiastic about this trip, it may be of little interest to others.

A second commonality which appears to exist is the conclusion that ample evaluation models are available and adequate knowledge exists with regard to their implementation (Leary, 1982). From the earliest model (Sandefur, 1970) to those presented at the spring symposium (Craig, 1982; Galluzzo, 1982), it is evident that what we know about evaluation can be translated into a multitude of evaluation models. However, it remains unclear as to whether the assumptions underlying program evaluation models are as well delineated as those of the well-documented research designs in education. In educational research, it is much easier to recognize that when given certain conditions a particular design is most appropriate. In the area of program evaluation, it does not appear clear that, for example, model A, effectively implemented at X, can be implemented at Y, or perhaps more importantly, if it is even appropriate at Y. Whether or not a "generic" model exists which is applicable in the vast majority of settings or if, in reality, models or designs must be institutionally specific, remains an important question. Even if we concur on where we are going, appropriate bait, tackle, and equipment remains debatable.

Both Borich and Stufflebeam have stressed another element on which it appears that agreement exists; namely, teacher preparation programs are implemented by people within varying program structures. Thus, the alternation of programs or structures ultimately means change for individuals. While the process of change has been researched and described (Hall, George & Rutherford, 1977) and while the change process in institutions of higher education (IHE) appears to be similar (Hall, Loucks, George, Lawrence, Sharp, & Schmid, 1978), it is much less clear what change strategies or interventions are effective in IHE's. While Borich has described the need to develop program "ownership" and Stufflebeam referred to the "institutionalization" of evaluation, both stressed the importance of significantly involving people in the change process. Thus, while it is recognized that both external and internal pressures exist which both promote and restrain program evaluation and change, the internal resistance to both evaluation and change may be far more important than has been previously addressed.

Finally, as noted by virtually all of the authors, but particularly evident in Smith's paper, the evaluation of teacher education programs must address multiple variables in multiple settings which may have multiple purposes. When one begins by attempting to define a teacher education program (T.E.P.), the complexity of the task, in and of itself, is enormous. In any given institution consideration must be given to identifying if a single T.E.P. program exists, if several exist, or if a generic core exists upon which T.E.P.'s are in fact a series of differentiated outcomes, a kind of branching effect. In large institutions, the question of "what is the program" becomes even more complex with multiple sections of courses taught by multiple faculty members. Thus, while students may take the same courses, it still may remain debatable that they have in fact been through the same preparation program.

Given these conclusions, future directions for evaluation in teacher education programs appear to be blocked or at the least limited. While many models exist, a wealth of evaluation instruments and approaches are available, and interested professionals perceive a need to proceed, the most critical element, namely the professional education faculty in the IHE's, has yet to be convinced of the value and

importance of program evaluation. Be it an argument of program ownership, inconvenience, or academic freedom, it remains that few IHE's internally provide meaningful rewards for curriculum development or program evaluation and change.

Given this admittedly bleak picture, one alternative might be that it is time to abandon this whole idea of a group fishing trip and simply go upon separate vacations. However, I think at this point that would be unwise. It appears highly unlikely that pressures for evaluation will dissipate. Public concern for the quality of education certainly appears to be increasing rather than decreasing. Demands for accountability of preparation programs from state certification agencies and professional organizations are becoming more prevalent and stringent. In turn, institutions responsible for teacher preparation have begun to recognize the need for greater and more systematic attention to be paid to program development and evaluation efforts.

Given that schools need and will increasingly need more effective teachers, then more effective preparation programs will also be needed. Program evaluation efforts may be able to provide the impetus for the development of those effective preparation programs. Thus, I would propose intensifying rather than abandoning program evaluation efforts and considering several alternative directions.

Future Directions

Teacher educators, particularly those interested in program evaluation, need to continue the scholarly debate and communication regarding what is to be evaluated, why programs should be evaluated, and who should be responsible for and have input into those evaluation efforts. However, it would appear essential that others become involved with those questions. Particularly important is the need to communicate what is already known to those seeking information or those facing evaluation requirements. While it may be questionable that such debate or communication in and of itself will lead to the development of "ownership" or the "institutionalization" of evaluation efforts, the change literature would suggest a need to develop "awareness" among a larger audience. Increasing dissemination efforts appears to be

extremely important. With the uncertainty of the future of specifically focused dissemination centers, new mechanisms need to be considered for communication.

Beyond increasing awareness and dissemination efforts, a need also exists to provide assistance to those institutions actually involved or beginning to be involved in evaluation efforts. As previously noted, interested and experienced evaluators exist, but apparently in a limited number of institutions. Various alternatives might include evaluator training sessions at professional meetings or the development of a consultant cadre available for either institutional assistance or state/regional training workshops. While most professional educators have been trained in research design, few may possess program evaluation expertise. Consideration must be given to identifying various means by which educators may utilize or develop program evaluation skills.

As has been previously noted, it appears that general consensus exists that ample models are available and that the real question is not as much one of which model or design to use, but simply how to put a model in place which will ultimately impact programs. The need to develop "ownership" of both programs and evaluation is critical to the future direction of teacher education.

However, the numerous research efforts regarding the change process and change implementation have primarily examined public school education. Much less is known regarding change in higher education. The kinds of leadership styles, strategies, and institutional processes which promote or constrain change need serious study. Thus, consideration must be given toward increasing research and evaluation efforts which examine the change process in higher education.

Given that past research has provided substantial evidence regarding effective teachers, administrators, and schools, collaborative efforts to link research, preparation, and practice (Erly, 1982) need to be pursued. While certainly IHE faculty need to be involved with "real world" education, public school personnel need to develop a better understanding of what IHE's can and cannot provide in the preparation of educational personnel. It may well emerge that effective teacher preparation and program evaluation is best accomplished by public

schools and program preparation institutions having joint responsibility.

Developing linkages between preparation programs and practice is a difficult task. Practitioners, already facing multiple time demands, have more frequently than ever before been unwilling to be involved in the preparation of teachers. Program evaluation, if approached from a collaborative effort, may provide the vehicle for linkage. Thus, further efforts in collaborative evaluation need investigation. Consideration must be given to identifying the important characteristics, structures, and processes which enhance collaboration.

Schools and colleges of education have frequently held a less than highly prestigious position within IHE's. Even in those institutions which originated as teacher preparation schools and have subsequently moved to college or university status, institutional resource support for teacher education has often been very limited. During the past several decades educators were often faced with heavy teaching and service loads, with little time left for research or evaluation efforts.

With the decline in enrollments in teacher education programs, institutions have cut sharply into the already limited resources of schools and colleges of education. Thus, while many administrators verbalize the need for and importance of program evaluation efforts, it remains a low priority even within the teacher preparation program itself. As faculty have come under increasing pressures to pursue scholarly activity, institutional rewards have been more often directed toward research and writing than program development and evaluation. If program evaluation is to impact meaningfully on programs, administrators must make more than a verbal commitment to its value. Administrator, as well as faculty members, must develop "ownership" of program evaluation and "institutionalize" it. This may imply a need for specific individuals to be identified, held responsible for, and rewarded for program evaluation efforts. The identification of those administrative characteristics, commitments, and rewards which facilitate the "ownership" of programs and the "institutionalization" of evaluation need to be considered.

External pressures for evaluation, be they from state or professional accreditation agencies, have to date been the major impetus

for program evaluation efforts. Institutional response to standards for accreditation has been widely varied, often leading to attempts to "tighten standards." Because accreditation is valued more than systematic program evaluation itself, efforts have been directed toward simply meeting the standards rather than viewing the value of the evaluation information relative to the program itself.

Until either program evaluation experts reach consensus on the most appropriate model and approaches to use or accrediting agencies delineate the standards more explicitly, current practices will prevail. Accrediting agencies really hold a two-edged sword. If institutions do program evaluation and if that evaluation does indicate program weaknesses, they face loss of program accreditation. If institutions do not demonstrate program evaluation efforts, they do not meet the evaluation standards.

However, it would appear that another alternative is available. Accrediting agencies should be encouraged to support evaluation experimentation by institutions without jeopardizing the accreditation of the programs of those institutions. Institutions need strong encouragement to test the models developed, to utilize approaches previously tried elsewhere, and to work in collaborative efforts with other institutions in building a common data base for teacher education programs. Consideration must be given to identifying support mechanisms which accrediting agencies might use to promote program evaluation efforts.

In summary, the future of teacher education program evaluation cannot be focused in a single direction. Rather, it is essential that efforts be given to movement concurrently in several directions. It is apparent that, if program evaluation is to impact teacher preparation programs, it must be encouraged, supported, and rewarded by multiple audiences. It is also apparent that this encouragement and support will not develop without active movement by those currently interested and involved in program evaluation. The next task at hand is for individuals to determine what directions they are willing to pursue.

References

Craig, S. B., Jr. Experimental program evaluation models. Pennsylvania Department of Education, Office of Higher Education, April, 1982.

Erly, M. C. A practitioner's perceptions regarding problems in assessing the effectiveness of teacher education programs. A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Galluzzo, G. R. Program evaluation in teacher education: From admissions to follow-up. A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Hall, G. E., Loucks, S. F., George, A. A., Lawrence, J., Sharp, B., & Schmid, R. An exploratory study of implementation of dean's grant projects in the southeast. Austin: Research and Development Center for Teacher Education, The University of Texas, 1978.

Hall, G. E., George, A. A., & Rutherford, W. L. Measuring stages of concern about the innovation: A manual for use of the SoC questionnaire. Austin: Research and Development Center for Teacher Education, The University of Texas, 1977.

Leary, J. A personal perspective on an important symposium. Guest Editorial, TEPFU Newsletter, June, 1982.

Roth, R. A. Requirements of a data base for effective program utilization. A working paper of the Symposium on the Evaluation of Teacher Education, University of Texas, Austin, 1982.

Sandefur, J. T. An illustrated model for the evaluation of teacher education graduates. Washington, DC: AACTE, 1970.

APPENDIX

LIST OF PARTICIPANTS

April, 1982

AUSTIN CONFERENCE

LIST OF PARTICIPANTS--TEPFU II 1982

A key planner unable to attend:

Dean Corrigan — College of Ed. Texas A&M University, College Station, TX 77843

Name	Address	Telephone
Ronald D. Adams	Director of Educational Research Western Kentucky University Bowling Green, KY 42101	502/745-2197
Patricia Alexander	College of Education Texas A&M University College Station, TX 77843	713/845-8398
Jerry B. Ayers	P. O. Box 5116 Tennessee Technological University Cookeville, TN 38501	615/528-3177
William Bechtol	Department of Education Southwest Texas State University San Marcos, TX 78666	
Lowell J. Bethel	Department of Curriculum and Instruction College of Education Education Building 340D University of Texas Austin, TX 78712	512/471-3223
Gary D. Borich	Department of Educational Psychology College of Education Education Building 518 University of Texas Austin, TX 78712	512/471-1044
Oliver H. Bown	Research and Development Center for Teacher Education Education Annex 3.203 University of Texas Austin, TX 78712	512/471-1343
Caseel D. Burke	Weber State University 3750 Harrison Blvd. Ogden, UT 84408	801/626-6622
Jim Craig	Psychology Department College of Education 226 Western Kentucky University Bowling Green, KY 42101	
Sam B. Craig, Jr.	Special Asst. to Commissioner of Education P. O. Box 911 Pennsylvania Dept. of Education Bureau Harrisburg, PA 17126	717/787-7013

List of TEPFU II Participants Cont'd.:

Name	Address	Telephone
Jon Denton	College of Education Texas A&M University College Station, TX 77843	
Betty Dillon-Peterson	Director of Staff Development Lincoln Public Schools P. O. Box 82889 Lincoln, NE 68510	402/473-0271
Maurice Early	Prince George's County Schools Upper Marlboro, MD 20870	301/952-4306 -4307
Joe L. Frost	Department of Curriculum and Instruction College of Education Education Building 406 University of Texas Austin, TX 78712	512/471-5942
Gary Galuzzo	EPIC Glassboro State College Glassboro, NJ 08028	609/455-7318
William Gardner	104 Burton Hall 178 Pillsbury Dr., S.E. University of Minnesota Minneapolis, MN 55455	612/373-9880
Gene E. Hall	Research and Development Center for Teacher Education Education Annex 3.114 University of Texas Austin, TX 78712	512/471-3844
Don Herring	College of Education Texas A&M University College Station, TX 77843	713/845-2951
Shirley M. Hord	Research and Development Center for Teacher Education Education Annex 3.114 University of Texas Austin, TX 78712	512/471-3844
W. Robert Houston	College of Education University of Houston Cullen Blvd. Houston, TX 77004	713/749-3621
Leslie L. Huling	Research and Development Center for Teacher Education Education Annex 3.114 University of Texas Austin, TX 78712	512/471-3844

List of TEPFU II Participants Cont'd.:

Name	Address	Telephone
Ann D. Hungerman	1302 SEB University of Michigan Ann Arbor, MI 48109	313/763-0650
Charles E. Johnson	College of Education University of Georgia Athens, GA 30601	404/542-4244
Donald Kachur	Department of Curriculum and Instruction College of Education Illinois State University Normal, IL 61/61	309/452-8071
Lorrin G. Kennamer	Dean of Education College of Education Education Building 210 University of Texas Austin, TX 78712	512/471-7255
Al Kilgore	Department of Secondary Education 215 Henzlik Hall University of Nebraska Lincoln, NE 68588	402/472-2015
Jim Leary	Frontier School Division 1402 Notre Dame Ave. Winnipeg, Manitoba CANADA R3E 3G5	
Douglas Palmer	College of Education Texas A&M University College Station, TX 77843	
William Peters	College of Education Texas A&M University College Station, TX 77843	713/845-8360
James Raths	College of Education University of Illinois 1210 S. 6th St. Champaign, IL 61820	
Maynard Reynolds	Department of Psycho-educational Studies 249 Burton Hall 178 Pillsbury Dr. S.E. University of Minnesota Minneapolis, MN 55455	612/373-4854

List of TEPFU II Participants Cont'd.:

Name	Address	Telephone
Robert A. Roth	904 Canton Drive Lansing, MI 48917	
William L. Rutherford	Research and Development Center for Teacher Education Education Annex 3.114 University of Texas Austin, TX 78712	512/471-3844
J. T. Sandefur	Dean College of Education Western Kentucky University Bowling Green, KY 42101	
Tom Savage	College of Education Texas A&M University College Station, TX 77843	
Nick Smith	Evaluation, Research and Assessment Division Northwest Regional Educational Laboratory 300 S. W. Sixth Avenue Portland, OR 97204	503/248-6800 Ext. 384
Bob Soar	University of Florida Gainesville, FL 32611	
Russ Spillman	Office of Program Development The Ohio State University 163 Arps Hall 1945 North High Street Columbus, OH 43210	
Marlene Strathe	Director of Educational Development and Research Center University of Northern Iowa Cedar Falls, IA 50613	319/273-6052
Daniel Stufflebeam	6660 Ravine Road Kalamazoo, MI 49009	
Edward White	Director, EPIC Program Glassboro State College Glassboro, NJ 08028	609/445-7318 -7319
Nancy L. Zimpher	The Ohio State University 163 Arps Hall 1945 North High Street Columbus, OH 43210	